

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日            2 0 0 4 年   2 月   3 日  
Date of Application:

出 願 番 号            特 願 2 0 0 4 - 0 2 6 3 5 6  
Application Number:  
[ST. 10/C]:            [ J P 2 0 0 4 - 0 2 6 3 5 6 ]

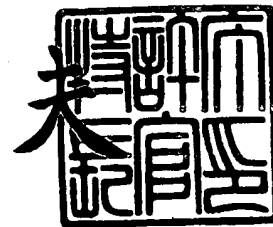
出 願 人            株式会社日立製作所  
Applicant(s):

CERTIFIED COPY OF  
PRIORITY

2 0 0 4 年   2 月 1 9 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



CERTIFIED COPY OF  
DOCUMENT

出証番号   出証特 2 0 0 4 - 3 0 1 1 3 1 1

CERTIFIED COPY OF  
PRIORITY DOCUMENT

【書類名】 特許願  
【整理番号】 NT03P0796  
【提出日】 平成16年 2月 3日  
【あて先】 特許庁長官 殿  
【国際特許分類】 G06F 12/00  
【発明者】  
    【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所  
                                システム開発研究所内  
    【氏名】 江口 賢哲  
【発明者】  
    【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所  
                                システム開発研究所内  
    【氏名】 山本 康友  
【発明者】  
    【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所  
                                システム開発研究所内  
    【氏名】 荒川 敬史  
【発明者】  
    【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R  
                                A I Dシステム事業部内  
    【氏名】 平川 裕介  
【特許出願人】  
    【識別番号】 000005108  
    【氏名又は名称】 株式会社日立製作所  
【代理人】  
    【識別番号】 100068504  
    【弁理士】  
    【氏名又は名称】 小川 勝男  
    【電話番号】 03-3537-1621  
【選任した代理人】  
    【識別番号】 100086656  
    【弁理士】  
    【氏名又は名称】 田中 恭助  
    【電話番号】 03-3537-1621  
【選任した代理人】  
    【識別番号】 100094352  
    【弁理士】  
    【氏名又は名称】 佐々木 孝  
    【電話番号】 03-3537-1621  
【手数料の表示】  
    【予納台帳番号】 081423  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1

**【書類名】 特許請求の範囲****【請求項 1】**

ホストと通信路を介して接続される第 1 の記憶装置システム及び第 2 の記憶装置システムを含み、第 2 の記憶装置システムは該第 1 の記憶装置システムから送られるデータのコピーを記憶するストレージサブシステムにおける、該第 2 の記憶装置システムに記憶されたデータを復元するデータの復元方法であって、

第 1 の記憶装置システムは、該ホストからの入出力要求を処理し、かつ第 2 の記憶装置システムに対して入出力処理の結果、更新されたデータを送信し、

第 2 の記憶装置システムは、第 1 の記憶装置システムからの受信したデータを更新ログデータとして保存し、

ホストはアプリケーションの状態確定するコマンドをデータとして第 1 の記憶装置システムに送信し、第 1 の記憶装置システムは該データを第 2 の記憶装置システムに送信し、

かつ該ホストと第 2 の記憶装置システムは、該コマンドに対応した識別子を双方で保持し、該識別子とログデータとを関連付けることにより、該ホストが任意の時点で該識別子を指示することによって第 2 の記憶装置システムで任意の時点のデータを復元することを特徴とするデータの復元方法。

**【請求項 2】**

前記ホストは、リモートサイトにある該第 2 の記憶装置システムに識別子の入出力指示を発行することを特徴とする請求項 1 のデータの復元方法。

**【請求項 3】**

リモートサイトにある該第 2 の記憶装置システムは、ホストの識別子の入出力指示を受領し、データの更新ログと該識別子とを関連つけて記憶装置に記憶すること特徴とする請求項 1 のデータの復元方法。

**【請求項 4】**

第 2 の記憶装置システムに記憶されているデータを復元する場合、該ホストから送信され、受信した識別子と一致する識別子を検索し、対象とする識別子を探したらコピー先の記憶装置に格納されたデータと、一致した識別子と関連付けられたログデータより前に記録されたログデータの内容を用いて、コピー元の記憶装置にデータを復元することを特徴とする請求項 1 のデータの復元方法。

**【請求項 5】**

該第 1 の記憶装置システムは、ホストからログデータの取得開始コマンド、及び該記憶装置のペア状態を中断するコマンドを受領すると、ペア関係にある第 2 の記憶装置システムにある記憶装置を確認して、ペアの状態をサスペンドすることを特徴とする請求項 1 乃至 4 のいずれかのデータの復元方法。

**【請求項 6】**

該第 2 の記憶装置システムは、該ホストで発行されたマークコマンドを受領すると、ログデータを取得する記憶装置を確認し、かつ

取得されたログデータに対してマーク ID 及びタイマ値を含むマークデータの対応を設定することを特徴とする請求項 1 乃至 5 のいずれかのデータの復元方法。

**【請求項 7】**

ホストとネットワークによって接続された記憶装置を備えた記憶装置システムを含む計算機システムにおける該記憶装置に記憶されるデータの処理方法において、

ホストは、記憶装置システムに対して記憶装置に格納されているデータのコピーの作成、保存を要求するステップと、ホストの処理によるデータの更新部分の記録を要求するステップと、計算機システムのある時点の状態を識別する識別情報を記憶装置システムに送信するステップと、

該記憶装置システムは、ホストの要求に回答して、該記憶装置のデータのコピーを作成して保存するステップと、該記憶装置の内容が更新されたときに、更新前後のデータ及び更新場所を示す情報をログデータとして保存するステップと、ホストより送信される識別情報を保持するステップと、該ログデータと識別情報を関連付けるステップと、

を有することを特徴とするデータ処理方法。

【請求項 8】

該記憶装置に記憶された内容がある時点の状態に復旧する場合、ホストは状態識別情報を指定してデータの復旧要求を記憶装置システムに送信するステップと、  
該記憶装置システムは、受信された該識別情報を判別し、該データのコピーとログデータを用いてデータをリストアするステップと、  
を更に有することと特徴とする請求項 7 記載のデータの復元方法。

【請求項 9】

該ホスト及び記憶装置システムで共有される該識別情報を、該記憶装置システムで識別情報と更新履歴を関連付けて管理し、該ホストからの指示に応じて、特定の識別情報で示される更新履歴までの該記憶装置に格納されたデータを復元することを特徴とする請求項 8 記載のデータの復元方法。

【請求項 10】

前記ログデータは少なくとも、特定の識別情報であるか否かを示すマークフラグのためのエントリと、ログの識別情報のためのエントリと、該ログを取得した時間を示すタイマ値のためのエントリと、データ領域となるエントリとを含み、  
該マークフラグが特定の識別情報を示す場合、該データ領域のエントリは、少なくともマークの識別情報のエントリと、マークを付与した時間を示すタイマ値のためのエントリを表すエントリを規定することを特徴とする請求項 8 又は 9 記載のデータの復元方法。

【請求項 11】

該ホストは、プログラム実行に際してファイルをクローズする時、又はファイルをセーブする時に、該ログデータと識別情報を関連付けるための特定のコマンドを生成して該記憶装置システムに送信し、該記憶装置システムで該コマンドを実行して、該マークフラグを特定の状態にセットし、かつ前記エントリにマークの識別情報とタイマ値を格納したログデータを記憶装置に保存することを特徴とする請求項 10 記載のデータの復元方法。

【請求項 12】

ホストと通信路を介して接続される第 1 の記憶装置システム及び第 2 の記憶装置システムを含み、該第 1 の記憶装置システムに記憶するデータのコピーを第 2 の記憶装置システムに記憶するストレージサブシステムにおいて、

該第 1 の記憶装置システムは；

複数の論理記憶装置を有する記憶装置と、

該記憶装置に入力又は出力されるデータを一時的に格納するキャッシュメモリと、

少なくとも該論理記憶装置に関する管理情報、該第 1 の記憶装置システムと第 2 の記憶装置システムのペア関係の構成を規定する管理情報、および該ホストからのコマンドを処理するプログラムを格納するメモリと、

該プログラムを実行するプロセッサと、を有し、

該第 2 の記憶装置システムは；

複数の論理記憶装置を有し、その内ある論理記憶装置はペアを構成する該第 1 の記憶装置システムの論理記憶装置に記憶されるデータのコピーを格納し、他のある論理記憶装置は該第 1 の記憶装置システムで生成されたログデータを記憶するために割当てられる記憶装置と、

該記憶装置に入力又は出力されるデータを一時的に格納するキャッシュメモリと、

少なくとも該論理記憶装置に関する管理情報、該第 1 の記憶装置システムと第 2 の記憶装置システムのペア関係の構成を規定する管理情報、ログの管理情報、および該ホストからのコマンドを処理するプログラムを格納するメモリと、

該プログラムを実行するプロセッサと、を有し、

該第 1 の記憶装置システムのある論理記憶装置の内容が更新された場合、更新されたデータ及び更新場所を示す情報をログデータとして該第 2 の記憶装置システムに送信して、該論理記憶装置に格納し、かつ該該ホストから送信された識別情報と該ログデータを対応付ける管理情報を該メモリに格納することを有することを特徴とするストレージサブシ

テム。

【請求項 13】

該第2の記憶装置システムにおける該論理記憶装置に格納されたログデータをある時点の状態に復旧する場合、該第2の記憶装置システムは、ホストから送信されたデータの復旧要求を受信して、該識別情報に関して該メモリに格納された前記管理情報を参照して該論理記憶装置に格納されたログデータをリストアすることを特徴とする請求項12記載のデータストレージサブシステム。

【請求項 14】

該第1の記憶装置システムの該メモリ及び該第2の記憶装置システムの該メモリは、該ホストが識別できる該記憶装置のアドレスと、該記憶装置内の論理アドレスの対応を登録する記憶装置管理情報テーブルと、該記憶装置内の論理アドレスと、該論理記憶装置が配置されているRAIDグループに関するアドレスと、該RAIDグループを形成するディスクに関するアドレスの対応を登録する記憶装置管理情報テーブルと、論理記憶装置の番号に対応してボリューム構成情報を登録するボリューム構成情報テーブルと、ペアを構成する該第1の記憶装置システム内の論理記憶装置の番号と該第2の記憶装置システム内の論理記憶装置番号の対応を登録するペア管理情報テーブルと、を有することを特徴とする請求項12又は13記載のデータストレージサブシステム。

【請求項 15】

少なくとも該第2の記憶装置システムは、ログボリュームグループごとにログボリュームグループ内の論理記憶装置に関する管理情報を登録するログボリュームグループ情報テーブルと、ログデータを該論理記憶装置に格納されるときに付与されるログIDと、ログデータが取得された時のタイマ値を対応付けて管理情報として登録するログID管理情報テーブルと、ログデータが該論理記憶装置に格納されるときに付与される、ホストから送信された識別情報と、ログデータが格納される時のタイマ値を対応付けて管理情報として登録するマークID管理情報テーブルと、を有することを特徴とする請求項14記載のデータストレージサブシステム。

【請求項 16】

前記第1の記憶装置システムにおいて、前記コマンド処理プログラムは、該ホストから送信されるコマンドを処理し、該コマンドがマークコマンドの場合には、ログデータを作成して、該識別情報を付与するための処理を行い、I/O処理コマンドの場合には、該キャッシュメモリのヒット又はミスヒットを判定し、ライトデータを該キャッシュメモリに書き込み、又は該キャッシュメモリからリードデータを読み出す処理を行うことを特徴とする請求項12乃至15のいずれか記載のデータストレージサブシステム。

【請求項 17】

該第2の記憶装置システムは、該ホストで発行されたマークコマンドを受領すると、ログデータを取得する記憶装置を確認し、かつ取得されたログデータに対してマークID及びタイマ値を含むマークデータの対応を設定することを特徴とする請求項12乃至16のいずれか記載のデータストレージサブシステム。

【請求項 18】

ホストと通信路を介して接続される該第1の記憶装置システムに記憶されるデータのコピーを第2の記憶装置システムにおいて、複数の論理記憶装置を有し、その内ある論理記憶装置はペアを構成する該第1の記憶装置システムの論理記憶装置に記憶されるデータのコピーを格納し、他のある論理記憶装置は該第1の記憶装置システムで生成されたログデータを記憶するために割当てられる記憶装

置と、

該記憶装置に入力又は出力されるデータを一時的に格納するキャッシュメモリと、  
少なくとも該論理記憶装置に関する管理情報、該第1の記憶装置システムと第2の記憶装置システムのペア関係の構成を規定する管理情報、ログの管理情報、および該ホストからのコマンドを処理するプログラムを格納するメモリと、

該プログラムを実行するプロセッサと、を有し、

該第1の記憶装置システムのある論理記憶装置の内容が更新された場合、更新されたデータ及び更新場所を示す情報をログデータとして該第2の記憶装置システムに送信して、該論理記憶装置に格納し、かつ該該ホストから送信された識別情報と該ログデータを対応付ける管理情報を該メモリに格納することを有することを特徴とする記憶装置システム。

【請求項19】

該論理記憶装置に格納されたログデータがある時点の状態に復旧する場合、該第2の記憶装置システムは、ホストから送信されたデータの復旧要求を受信して、該識別情報に関して該メモリに格納された前記管理情報を参照して該論理記憶装置に格納されたログデータをリストアすることを特徴とする請求項18記載の記憶装置システム。

【請求項20】

ホストと通信路を介して接続される第1の記憶装置システム及び第2の記憶装置システムを含み、該第1の記憶装置システムに記憶するデータのコピーを第2の記憶装置システムに記憶するストレージサブシステムにおいて、

該第1の記憶装置システムは；

複数の論理記憶装置を有する記憶装置と、

該記憶装置に入力又は出力されるデータを一時的に格納するキャッシュメモリと、

少なくとも該論理記憶装置に関する管理情報、該第1の記憶装置システムと第2の記憶装置システムのペア関係の構成を規定する管理情報、および該ホストからのコマンドを処理するプログラムを格納するメモリと、

該プログラムを実行するプロセッサと、を有し、

前記コマンド処理プログラムは、該ホストから送信されるコマンドを処理し、該コマンドがマークコマンドの場合には、ログデータを作成して、該識別情報を付与するための処理を行い、I/O処理コマンドの場合には、該キャッシュメモリのヒット又はミスヒットを判定し、ライトデータを該キャッシュメモリに書き込み、又は該キャッシュメモリからリードデータを読み出す処理を行い、

該第2の記憶装置システムは；

複数の論理記憶装置を有し、その内ある論理記憶装置はペアを構成する該第1の記憶装置システムの論理記憶装置に記憶されるデータのコピーを格納し、他のある論理記憶装置は該第1の記憶装置システムで生成されたログデータを記憶するために割当てられる記憶装置と、

該記憶装置に入力又は出力されるデータを一時的に格納するキャッシュメモリと、

少なくとも該論理記憶装置に関する管理情報、該第1の記憶装置システムと第2の記憶装置システムのペア関係の構成を規定する管理情報、ログの管理情報、および該ホストからのコマンドを処理するプログラムを格納するメモリと、

該プログラムを実行するプロセッサと、を有し、

該第1の記憶装置システムのある論理記憶装置の内容が更新された場合、更新されたデータ及び更新場所を示す情報をログデータとして該第2の記憶装置システムに送信して、該論理記憶装置に格納し、かつ該該ホストから送信された識別情報と該ログデータを対応付ける管理情報を該メモリに格納し、

該第1の記憶装置システムは、該ホストからログデータの取得開始コマンド、及び該記憶装置のペア状態を中断するコマンドを受領すると、ペア関係にある第2の記憶装置システムにある記憶装置を確認して、ペアの状態をサスペンドし、

該第2の記憶装置システムにおける該論理記憶装置に格納されたログデータがある時点の状態に復旧する場合、該第2の記憶装置システムは、ホストから送信されたデータの復旧

● 要求を受信して、該識別情報に関して該メモリに格納された前記管理情報を参照して該論理記憶装置に格納されたログデータをリストアすることを特徴とするデータストレージサブシステム。

**【書類名】明細書****【発明の名称】ストレージサブシステム****【技術分野】****【0001】**

本発明は、ストレージサブシステムに係り、特にライト I/O ログを取得するストレージサブシステムの遠隔コピー、及び災害時におけるログの復旧方法に関する。

**【背景技術】****【0002】**

計算機と記憶装置システムとをネットワークにより接続し、計算機で処理されるデータを、ネットワークを介して送受信して記憶装置システムに格納する計算機システムが実用化されている。計算機システムで実行されるオンライン処理やバッチ処理において、プログラムのバグや記憶装置システムの障害などによってこれらの処理が異常終了し、記憶装置システムに格納されたデータに矛盾が生じる事態になることがある。また、人為的ミスによって記憶装置システムに格納されたデータが消去されてしまうこともある。

**【0003】**

このような状態になった計算機システムのデータを回復させるために、データの矛盾を解消して途中で止まった処理を再開させたり、あるいは途中で止まった処理をもう一度実行し直したりするための技術の一つとして、データのバックアップとリストアによるデータ回復技術が知られている。

**【0004】**

バックアップおよびリストアに関する従来技術として、例えば米国特許番号 5, 263, 154 号公報（特許文献 1）には、ユーザが指定した時点における記憶装置システムに格納されたデータを、記憶装置システムに接続されたホストからのデータの入出力（以下、「I/O」）を止めることなく磁気テープに複製し（以下、「データのバックアップ」）、その複製されたデータ（以下、「バックアップデータ」）を用いてデータの回復（以下、「リストア」）する技術が開示されている。

**【0005】**

また、特開 2001-216185 号公報（特許文献 2）には、データのリストアにかかる時間を短縮するために、データのバックアップが実行された後、データが更新された場所における情報を差分情報として保持し、記憶装置システムに格納されたデータをバックアップデータでリストアする際に、バックアップデータのうち、差分情報で示されるデータの部分のみをデータのリストアに用いる技術が開示されている。

**【0006】**

また、米国特許番号 5, 544, 347 号公報（特許文献 3）、米国特許番号 5, 742, 792 号公報（特許文献 4）には、リモートにある記憶装置システムで、ホストに独立に、データをコピーするという技術が開示されている。この技術によれば、ある業務サイトにてホストによって作成された、記憶装置システム内の業務ボリュームのコピーをリモートサイトの記憶装置システム内のボリュームに作成することができる。そのため、業務サイトが天災やテロなどの被災に会い、計算機システムが障害を起こし、業務が継続できなくなった場合を考慮して、リモートサイトにおいて業務の復旧のためのデータが用意することができる。

**【0007】**

【特許文献 1】 米国特許番号 5, 263, 154 号公報

**【0008】**

【特許文献 2】 特開 2001-216185 号公報

【特許文献 3】 米国特許番号 5, 544, 347 号公報

【特許文献 4】 米国特許番号 5, 742, 792 号公報

**【発明の開示】****【発明が解決しようとする課題】****【0009】**



特許文献1に記載されたリストア処理では、磁気テープからバックアップデータを読み出す際、バックアップデータを取得した時点から更新されていない部分（即ち、記憶装置システムのデータと磁気テープのデータの内容が一致している部分）も磁気テープから読み出され、記憶装置システムに書き込まれる。このようなデータの転送は、無駄が多く、リストアのための時間が長くなる。

【0010】

特許文献2に開示されている技術では、特許文献1の技術に比べ、重複したデータの読み出しが発生しない分、リストアに係る時間は少なくなる。しかし、双方の技術をもってしても、データのバックアップの後から記憶装置システムが故障するまでの間に更新されたデータについては、データのリストアを行うことができない。データのバックアップ後に更新されたデータまでリストアしようとする、そのデータの更新の内容等をホスト側がログ等で管理する必要がある。このためホストへの負荷が大きく、またその処理に長い時間がかかる。

【0011】

また、特許文献3や特許文献4には、業務サイトが被災し、計算機システムが障害を起こして業務が継続できなくなった場合を想定して、リモートサイトで復旧のためのデータを作成する技術が開示されているが、災害中の書き込みにより生じる不正な書き込みに関するリモートサイトでのコピーの危険性に対する対策案については考慮されていない。

【0012】

本発明の目的は、障害発生前までの任意の時点におけるデータのリストア処理を高速に行う計算機システムを提供することにある。

【0013】

本発明の目的は、記憶装置システムに格納されたデータを復旧する場合に、ホストに負担をかけず、リモートサイトでデータをリストアする記憶装置システムを提供することにある。

【課題を解決するための手段】

【0014】

本発明は、好ましくは、ホストと通信路を介して接続される第1の記憶装置システム及び第2の記憶装置システムを含み、第2の記憶装置システムは第1の記憶装置システムから送られるデータのコピーを記憶するストレージサブシステムにおいて実現される。

このストレージサブシステムにおいて、第1の記憶装置システムは、ホストからの入出力要求を処理し、かつ第2の記憶装置システムに対して入出力処理の結果、更新されたデータを送信し、第2の記憶装置システムは、第1の記憶装置システムからの受信したデータを更新ログデータとして保存し、ホストはアプリケーションの状態確定するコマンドをデータとして第1の記憶装置システムに送信し、第1の記憶装置システムはそのデータを第2の記憶装置システムに送信し、かつホストと第2の記憶装置システムは、コマンドに対応した識別子を双方で保持し、その識別子とログデータとを関連付ける。データを復元する場合に、ホストは任意の時点で識別子を指示することによって第2の記憶装置システムで任意の時点のデータを復元する。

好ましい例では、ホストは、リモートサイトにある第2の記憶装置システムに状態識別情報の入出力指示を発行する。リモートサイトにある第2の記憶装置システムは、ホストの状態識別情報の入出力指示を受領し、データの更新ログと識別情報を関連つけて記憶装置に記憶する。

【発明の効果】

【0015】

本発明によれば、記憶装置システムに格納されたデータを復旧する場合に、ホストに負担をかけず、リモートサイトで短時間にデータを所定の状態までリストアすることができる。

【発明を実施するための最良の形態】

【0016】

以下、図面を用いて、本発明の一実施形態について説明する。

図1は、一実施形態による計算機システムのブロック図を示す。

この計算機システムは、第1のホスト計算機1（以下、ホスト1という）と、第2のホスト計算機2（以下、ホスト2という）と、ホスト1に通信路5を介して接続される第1の記憶装置システム3（以下単に記憶装置システム3という）と、ホスト2に通信路7を介して接続されると共に、記憶装置システム3と通信路6を介して接続される第2の記憶装置システム4（以下単に記憶装置システム4という）と、記憶装置システム3に通信路11を介して接続される第1の記憶装置システム管理装置8と、記憶装置システム4に通信路12を介して接続される第2の記憶装置システム管理装置（以下、管理装置という）9と、これらの構成装置と通信路10を介して接続される計算機システム管理装置13を含んで構成される。記憶装置システム3、4は、例えばディスク装置或いはディスクアレイのような記憶装置21、22、23及びそれらの制御装置を有して構成される。ここで第2系の側はリモートサイトとして位置付けされる。

#### 【0017】

ホスト1及びホスト2は、パーソナルコンピュータ、ワークステーション、メインフレーム等の計算機である。ホスト1では、その計算機の種類に応じたオペレーティングシステム（OS）や、様々な業務や用途に対応したアプリケーションプログラム（AP）例えばデータベース（DB）プログラムのような色々なプログラムが実行される。なおこの例では、簡単のため、ホスト1とホスト2とは1つずつ示されているが、通信路5または通信路7に接続されるホスト1またはホスト2は複数であってよい。

ホスト1は、計算機システムにおける所定の処理を実行する。即ち、ホスト1は情報処理に係る入出力処理に必要なコマンド及びデータを、記憶装置システム3との間で通信路5を用いて通信し、ホスト1で作成、変更したデータを記憶装置システム3に対してデータのライト要求を行い、また計算機処理に必要なデータのリード要求を行う。

#### 【0018】

記憶装置システム3は、通信路5を介して送信されるコマンドやデータを受信し、所定の処理を行い、ホスト1に対する所定の応答を実行する。記憶装置システム3と記憶装置システム4は、通信路6を介してそれらの間でコマンド及びデータの通信を行う。

記憶装置システム4の構成も記憶装置システム3と実質的に同じであるが、図1では複数の記憶装置22、23を有する点が相違する。詳しくは後述するが、2つある記憶装置の内一方の記憶装置22は記憶装置21のミラー用であり、他方の記憶装置23はログデータ記憶用に使用される。

#### 【0019】

管理装置8は、記憶装置システム3の設定を行う他、記憶装置システム3の障害、保守、構成、性能情報等の管理を行うための計算機である。同様に、管理装置9は、記憶装置システム4の設定を行うと共に、記憶装置システム3の障害、保守、構成、性能情報等の管理を行うための計算機である。例えば、計算機システムの管理者が、記憶装置システム3及び記憶装置システム4に、例えばボリュームのような論理的な記憶装置を設定する場合、データをバックアップするための記憶領域を設定する場合、又はデータを複製する際の記憶領域の対を設定する場合に、これらの管理装置8、9が用いられる。

システム管理者は、記憶装置システム3の保守・管理、記憶装置システム3が有する物理記憶装置の設定、及び記憶装置システム3と接続されるホスト1の設定等を行う場合に、管理装置8に設定したい内容を入力する。システム管理者が管理装置8に入力した内容は、通信路11、10を介して記憶装置システム3及びホスト1に送信される。

計算機システム管理装置13は、計算機システム全体の保守、管理を行うためのものであり、通常システムエンジニアにより操作され利用される。

#### 【0020】

通信路5及び7は、ホスト1、ホスト2が夫々記憶装置システム3、記憶装置システム4へI/Oの処理要求等を伝送するために使用される。通信路6は、記憶装置システム3と記憶装置システム4との間のI/Oの処理要求等を伝送するために使用される。通信路

5, 6, 7には、光ケーブルや銅線等が用いられる。これらの通信路5, 6, 7で利用される通信プロトコルとしては、例えばイーサネット（登録商標）、FDDI、ファイバチャネル、SCSI、Infiniband、TCP/IP、iSCSIなどがある。

通信路11及び通信路12は、記憶装置システム3や記憶装置システム4が、自身の障害、保守、構成、性能等の管理情報を管理装置8、9との間で転送するために使用される。通信路10は、ホスト1、2が管理装置8、9から管理情報を取得する場合にコマンドを転送したり、及びに記憶装置システム3、4の障害、保守、構成、性能等の管理情報を管理装置8、9からホスト1に送信したり使用される。通信路10, 11, 12で利用されるケーブル及び通信プロトコルは、通信路5, 6, 7と同じでもよいし、異なってもよい。

#### 【0021】

図2は、記憶装置システム3、4の構成を示す図である。

記憶装置システム3は、記憶装置システム制御装置101と記憶装置21から構成され、記憶装置21の記憶領域はホスト1からの入出力のために提供される。即ち記憶装置システム3は、ホスト1が使用するデータやプログラムを格納し、ホスト1のI/O処理要求を受信し、I/O処理要求に対応した処理を行い、その結果を所定のホスト1に送信する。

#### 【0022】

記憶制御装置101は、記憶装置21と接続される記憶装置アダプタ108、所定のプログラムを実行するプロセッサ105、プロセッサ105で実行されるプログラム、プログラムが動作する上で必要な情報、記憶装置システム3の設定情報及び構成情報等が格納される不揮発性制御メモリ107、この記憶装置システム3を通信路11と接続するためのネットワークアダプタ102、及び通信路5と接続するためのホストアダプタ103、通信路6と接続するためのリモートI/Oアダプタ104を有し、ホスト1からのI/O要求の処理及び記憶装置21の制御を行う。

#### 【0023】

記憶装置21は、ボリュームと称する複数の論理的な記憶装置109から構成され、夫々の記憶装置109にはユーザ用のデータが格納される。記憶装置21は、その記憶領域に障害が生じててもデータが損失しないように、冗長性を持つRAID (Redundant Arrays of Inexpensive Disks) 構成とするのが好ましい。記憶装置21の記憶媒体としては、例えば電氣的に不揮発な記憶媒体である磁気ディスクや不揮発性半導体メモリで構成されるシリコンディスク、光ディスク、光磁気ディスク又はハードディスク等の媒体が利用される。

尚、記憶制御装置101は、記憶装置システム3に複数存在してもよい。また記憶装置システム3の冗長性を確保するために、好ましくは、記憶制御装置101内の各構成要素への電源供給のための回路、記憶制御装置101内の各構成要素間でのデータの送受信を行う回路、キャッシュメモリ106、不揮発性制御メモリ107、記憶装置アダプタ108等を、それぞれ2重化された冗長構成とするのがよい。

#### 【0024】

不揮発性メモリ107は、記憶装置システム3内の処理を制御するために、コマンド処理プログラム111、リモートI/O処理プログラム112、コピーマネージャプログラム113、ペア管理プログラム114、ネットワーク処理プログラム119、記憶装置I/O処理プログラム121、システム構成管理プログラム122の各プログラムを格納する。不揮発性メモリ107は、またI/O処理のためにキャッシュメモリ106のある領域へアクセスを行うための排他処理用ビットや記憶装置109とキャッシュメモリ107との対応関係を示す情報を格納する。尚、不揮発性の制御メモリ107は、記憶制御装置101やプロセッサ105が複数存在する場合には、それらで共有するようにしてもよい。

キャッシュメモリ106は、ホスト1から記憶装置システム3に転送されるデータ又は記憶装置システム3からホスト1へ転送されるデータを一時的に格納する。

**【0025】**

ローカルエリアネットワーク120は、記憶制御装置101、キャッシュメモリ106、及び記憶装置109を相互に接続する。ローカルエリアネットワーク120は、例えば共有バス型の構成でもよいし、スター型等のネットワーク構成でもよい。

**【0026】**

記憶制御装置101は、不揮発性メモリ107に格納されたプログラムをプロセッサ104で実行することで、以下に説明する処理を制御する。コマンド処理プログラム111は、ホスト1からのI/O処理要求をホストアダプタ102で受信すると、受信したI/O処理要求の内容を解析する。解析の結果に従って、I/O処理要求の内容がデータの読み出しI/O（以下「リードI/O」）要求やデータの書き込みI/O（以下「ライトI/O」）処理要求等を実行する。ライトI/O処理要求の場合、記憶制御装置101は、ホスト1からのライトI/O処理要求に対する応答処理、例えば実際にホスト1から転送されるデータを受領できる状態にあるかどうかの応答処理を行い、更に転送されてくる更新用のデータ（以下「ライトデータ」）をキャッシュメモリ106又は記憶装置109の指定された場所へ書き込む。リードI/O処理要求の場合、記憶制御装置101は、リードI/O処理要求に対応するデータ（以下、「リードデータ」）を、キャッシュメモリ106もしくは記憶装置109の指定の場所から読み出してホスト1に転送する。またその他の処理、例えばホストとの入出力I/FがSCSIで、SCSIのInquiryコマンド（デバイスサーチを指示するコマンド）を受領した場合、ホストI/O処理プログラムは、コマンドで要求される処理内容に対応した動作の制御を行う。

**【0027】**

システム構成管理プログラム122は、記憶制御装置101が記憶装置システム3を管理する際に実行するプログラムであり、記憶装置管理情報115、116の作成、設定、変更、削除等を行う。記憶制御装置101は、システム構成管理プログラム122を実行することによって、管理装置8から入力される記憶装置109の定義やバックアップ/スナップショット用のペア対象となる記憶装置109の設定、LOG格納対象情報の登録等を行う。ここで、記憶装置管理情報115、116は、記憶装置システム内のアドレスとホスト記憶装置に対して入出力を行うためのアドレスとの対応関係を示すマッピング情報、ペア118に関する情報を保持する。

**【0028】**

記憶装置I/O処理プログラム121は、記憶制御装置101がキャッシュメモリ106又は記憶装置109に対してデータのリードライト処理を行う場合に実行されるプログラムである。リードライトI/O処理要求の実行に際しては、システム構成管理プログラム122を実行してアクセス先の記憶装置109の構成をチェックして、リードライトI/O処理要求が指定するリード又はライトデータが読み出され又は格納されるべき場所のアドレスがどの記憶装置109のアドレスに対応するかを計算し、その計算結果に基づいて、記憶装置109へのアクセスを行う。

**【0029】**

ペア管理プログラム114は、マスタボリューム（以下、「M-VOL」）とリモートボリューム（以下「R-VOL」）のペアを管理するためのプログラムである。記憶制御装置101は、ペア管理プログラム114を実行することで、ホスト1からの指示に従って、ある記憶装置109（「M-VOL」）及びM-VOL109に格納されたデータの複製を格納するリモートの記憶装置システム4内にある記憶装置22内の記憶装置（「R-VOL」）139について、ペア作成（Pair Create）、ペア中断（Pair Suspend）、ペア再開（Pair Resync）、ペア解除（Pair Delete）の処理を行う。

尚、1つのM-VOLに対して、複数のR-VOLを設定し、作成することもできる。また、R-VOLを新たなM-VOLとして、新たなM-VOLとペアになるR-VOLを設定、作成することもできる。

## 【0030】

ペア管理情報118は、あるM-VOL109がペア状態(Pair Duplex)、ペア作成状態(Pair Create)、ペア中断状態(Pair Suspend)にあるかどうかを示す情報を登録する。

## 【0031】

記憶制御装置101は、コピーマネージャプログラム113を実行することによって、ペア作成(Pair Create)時にM-VOL139の先頭アドレスから順次R-VOLにデータをコピーすることで、M-VOLに格納されたデータをR-VOLにバックアップする。さらに記憶制御装置101は、差分情報123を参照して、差異が有る部分のデータをM-VOLからR-VOLにコピーしたり、逆に、差分情報123を参照して、差異があるデータをR-VOLからM-VOLへコピーするようにリモートI/Os処理プログラム112に指示を出す。ここで差分情報123とは、M-VOLとR-VOLのコピーデータの不一致を防ぐために、記憶装置109の記憶領域に対応して設けられたビットマップであり、リモートコピーした後に記憶装置109の記憶領域のデータが書き換えられた場合、その更新された領域に対応してビットフラグを立てて管理する。また、ペア管理プログラム114は、ホスト1からのリストア要求に基づいて、指定された記憶装置109のデータをリストアする。尚、リストア処理の詳細は後述する。

## 【0032】

以上、マスタサイト側の記憶装置システム3の構成について述べた。一方、リモートサイト側の記憶装置システム4も、基本的には同様のプログラム及び情報を所有する。主な相違点を述べると、記憶装置22は、マスタサイトの記憶装置21のコピーデータを記憶するが、記憶装置23はマスタサイト側に発生したログデータを記憶する。もちろんマスタサイト側に発生したログデータは、マスタ側の記憶装置に記憶するようにしても良いが、図2の例では、そのログデータを通信路6を介してリモートの記憶装置システム4に送信して記憶装置23内の論理記憶装置140に格納するように構成される。尚、記憶装置22、23内の論理的記憶装置(ボリューム)139、140の構成は同様である。

## 【0033】

不揮発メモリ137に格納される情報に関して言うと、コマンド処理プログラム142、リモートI/O処理プログラム143、コピーマネージャプログラム144、ペア管理プログラム145、ネットワーク処理プログラム159、記憶装置I/O処理プログラム162、システム構成管理プログラム163の各プログラム、及び記憶装置管理情報150、151、ボリューム構成情報152、ペア管理情報153、差分情報164、等々は、マスタサイト側の対応するプログラム或いは情報と同様である。しかしながら、リモートサイトの記憶装置23にログデータ及びその管理情報を格納し、管理情報に基づいてログデータを復元処理するという特徴的な動作を行うために、不揮発メモリ137には、ログボリュームグループ管理プログラム146、使用量管理プログラム147、ログID管理プログラム148、Mark ID管理プログラム149、ログボリュームグループ構成情報154、ログボリュームグループ使用量管理情報155、ボリュームプール構成情報156、ログID管理情報157、Mark ID管理情報158が記憶される。これらのプログラム及び情報の機能、意味については、以後順次説明される。

## 【0034】

次に、図3を参照してホスト1の構成について説明する。ホスト1を例にして説明するが、ホスト2の構成も実質的にホスト1と同様である。ホスト1は、所定のプログラムを実行するプロセッサ201、プロセッサ201が実行するOSやAP及びAPが使用するデータを格納するために使用されるメモリ202、OSやAP、APが使用するデータが格納されるディスク装置207、通信路5を接続するI/Oアダプタ205、通信路10を接続するネットワークアダプタ206、フロッピー(登録商標)ディスク等の可搬記憶メディアからのデータの読み出し等を制御するリムーバ

ブル記憶ドライブ装置 209、液晶表示装置のような表示器 203、キーボード或いはマウスのような入力器 204、及びこれらの構成ユニットを接続し、OSやAPのデータや制御データの転送に用いられる Local I/O ネットワーク 208 とを有して構成される。

#### 【0035】

リムーバブル記憶ドライブ装置 209 で使用される可搬記憶媒体としては、CD-ROM、CD-R、CD-RW、DVD や MO 等の光ディスク、光磁気ディスクや、ハードディスクやフロッピーディスク等の磁気ディスク等がある。尚、以下に説明される各プログラムは、可搬記憶媒体からリムーバブル記憶ドライブ装置 209 を介して読み出されることで、あるいはネットワーク 4 又は 5 を経由することで、ホスト 1 のディスク装置 207 にインストールされる。尚、ホスト 1 は、冗長性確保のために、プロセッサ 201 等の構成ユニットを複数有していても良い。

#### 【0036】

次に、ホスト 1 で実行されるプログラムの例について説明する。これらのプログラム 210 は、ホスト 1 のディスク装置 207 又はメモリ 202 に格納され、プロセッサ 201 で実行される。ホスト 1 は、OS 230 の下で動作する AP 233、データベースマネジメントソフトウェア（以下「DBMS」）232 を有する。DBMS 232 は、OS 230、ファイルシステム（FS）231、ボリュームマネージャ（VM）235 等を介して記憶装置システム 3 にアクセスする。また、DBMS 232 は、ユーザが使用する他の AP 233 との間で、トランザクション処理等の I/O 処理のやり取りを行う。情報処理の性能を上げる目的で、ホスト 1 のメモリ 202 を用いてプロセッサ 201 が情報処理を実行する。

#### 【0037】

また、ホスト 1 が入出力を行う複数の論理記憶装置間で相関があるものをログボリュームグループ管理情報 248 として保持する。また、ペア管理情報 249 は、ホスト 1 が入出力を行う論理記憶装置である M-VOL と、それとペアになる R-VOL の管理情報を保持する。ペア操作プログラム 247 は、記憶装置システム 3 に対してペア操作（Pair Create（ペアの作成）、Pair Suspend（ペアの中断）、Pair Resync（ペアの再開）、Pair Delete（ペアの削除））を行う。ログ操作プログラム 246 は、ログの操作、例えばログ取得開始、ログ取得終了等を行う。ペア管理プログラム 245 は、記憶装置システム 3 のペアの状態を監視し、障害が無いか監視する。容量管理プログラム 244 は、記憶装置システム 3 の容量を情報として保持しログボリュームグループに含まれる論理記憶装置の全容量を算出する。マーク処理プログラム 243 は、AP 233 や DBMS 232 がファイルをセーブした場合や閉じた場合に、コミットした処理の後に呼び出され、マークデータを作成して記憶装置システム 3 に書き込む処理を行う。マーク ID 管理プログラム 242 は、マーク処理プログラム 243 がマークデータをライトする場合にタイマとマーク ID 情報 241 を用いてマークデータを作成する。マーク ID 情報 241 はログ ID に対する管理情報であり、マーク ID とログの時刻情報から成る。

#### 【0038】

次に、図 4、図 5 を参照して、記憶装置管理情報 115、150、及び記憶装置管理情報 116、151 について説明する。尚、符号は記憶装置システム 3 側を参照して説明する。

図 4 において、記憶装置管理情報 115 は、ホスト 1 に提供される記憶装置 21 に関するアドレスを登録するエントリ 301 と、記憶装置システム 3 で記憶装置 21 を統一的に識別するための論理的なアドレスを登録するエントリ 302 を有するテーブルである。

エントリ 301 は、ホスト計算機に提供される記憶装置の識別子を登録するエントリ 303、及びその内部アドレスを登録するエントリ 304 を有する。また、エントリ 302 は、記憶装置システム 3 で記憶装置を統一的に識別する記憶装置 21 内の論理記憶装置 109 の識別子 305 及びその内部アドレスを登録するエントリ 306 を有する。

## 【0039】

図5において、記憶装置管理情報116は、記憶装置システム3内で記憶装置21を统一的に識別する論理的アドレスを登録するエントリ401と、RAID Groupに関するアドレスを登録するエントリ402と、RAID Groupを構成するディスクに関するアドレスを登録するエントリ403を有するテーブルである。

更にエントリ401は、記憶装置システム3内で論理記憶装置109を统一的に識別する論理記憶装置番号を登録するエントリ404と、記憶装置に対応する内部アドレスを登録するエントリ405を有する。エントリ402は、記憶装置システム3内で論理記憶装置109を统一的に識別する論理記憶装置番号によって識別される論理記憶装置が配置されているRAID Groupを記憶装置システム3内で统一的に識別するRAID Group番号を登録するエントリ406と、RAID Groupを1つの記憶領域として扱った仮想空間の対応するアドレスを登録するエントリ407を有する。また、エントリ403は、RAID Groupを形成するディスクを記憶装置システム3内で统一的に識別するディスク番号を登録するエントリ407と、ディスク内部アドレスが登録されるエントリ408を有する。

## 【0040】

次に、図6を参照して、図2に示すVolume構成情報117、152について説明する。尚、Volume構成情報152も同様であるので、符号は記憶システム1側の符号を参照して記載する。

Volume構成情報117は記憶装置システム3におけるVolumeの構成に関する情報を登録するテーブルである。エントリ501には、記憶装置システム3内の論理記憶装置を统一的に扱う論理記憶装置番号が登録される。エントリ502には、ホストタイプが登録される。すなわち論理記憶装置に対して入出力を行うホストのOSが認識できる記憶装置のいずれに擬似化されている（エミュレートされる）かを示す情報、例えば、オープン系システムのOSが認識できる記憶装置であることを示す「OPEN」や、メインフレーム系のOSが認識できる記憶装置であることを示す「3990」等の情報が登録される。

## 【0041】

エントリ503には、ホスト1が入出力を行えるように入出力ポートに関連付けられているかどうかを示すパス定義情報が登録される。例えばI/OネットワークがFCだったら、論理記憶装置とFCのPortとの関連付けに関する情報が登録される。

エントリ504は、論理記憶装置の状態を登録するものであり、例えば論理記憶装置に障害が無く入出力が行えるノーマル状態（NORMAL）を示す情報や、障害などの理由で入出力が行えない状態（BLOCKED）を示す情報が登録される。更に障害情報としては、論理記憶装置が何らかの障害になったかどうかを示す情報が登録される。ここで、障害とは、主に論理記憶装置を構成する物理記憶装置の物理的障害や、管理者が意識的に記憶装置システムを閉塞状態にした場合等の論理的障害が含まれる。

## 【0042】

エントリ505はリザーブ情報を登録するものであり、例えば論理記憶装置が、R-VOLやログデータを格納するために予約されている状態にあるかを示す情報が登録される。リザーブ情報が登録されている論理記憶装置は、その他の用途、例えば新たに業務用論理記憶装置として割り当てなどが出来ない。

論理記憶装置109がペアを形成している場合には、ペア情報を登録するエントリ506にペア番号が登録され、記憶装置内でのログデータを取得する場合にはログボリュームグループ番号を登録するエントリ507にログボリュームグループ番号が登録される。エントリ506に、有効な情報例えばログボリューム番号が在れば、論理記憶装置がログ取得の対象、すなわちジャーナルモードの対象であるかどうかを示す情報が登録される。

またエントリ508には論理記憶装置109の容量が登録される。

## 【0043】

次に、図7を参照して、ペア管理情報118、153について説明する。尚、参照符号

はペア管理情報118側について示す。

ペア管理情報118は、ペアを構成する正副の論理記憶装置の識別子を登録するテーブルである。すなわち、エントリ601にはペア番号が登録される。エントリ602には、ペアを形成するマスタ側の記憶装置システム3内の統一的な論理記憶装置(M-VOL)番号が登録される。エントリ603には、M-VOLとペアを形成するリモート側の記憶装置システム4に在る論理記憶装置(R-VOL)を識別する情報が登録される。例えばエントリ603には、通信路6で接続される入出力用ポートの識別子とこのポートからアクセス可能な記憶装置システム4内で統一的な記憶装置の番号が登録される。エントリ604には、ペアの状態、例えばペア状態、ペア作成中、ペア中断中、ペア再形成中を示す情報が登録される。

#### 【0044】

次に、図8以降の図面を参照して、リモート側の記憶装置システム4側に特有なテーブルの構成について説明する。

図8は、ログボリュームグループ構成情報154のテーブルの構成を示す図である。ログボリュームグループ構成情報154は、ログボリュームグループごとに用意される。カラム701にはログボリュームグループ毎を識別する情報が格納される。カラム702には、ログボリュームグループにグルーピングされた論理記憶装置の総数が格納される。カラム703には、ログボリュームグループにグルーピングされた論理記憶装置の容量の総和が格納される。

#### 【0045】

カラム704には、ログボリュームグループにグルーピングされた論理記憶装置の識別子705と、HOST TYPE 706と、状態707と、容量708がログボリュームグループにグルーピングされた論理記憶装置の数分だけ登録される。カラム709には、ログボリュームグループで使用されるLog用Volumeの論理記憶装置の識別子の数が登録され、カラム710にはログボリュームグループで使用するLog用Volumeの論理記憶装置の容量の総和が登録される。

カラム710には、ログボリュームグループで使用するLog用Volumeの論理記憶装置の識別子711と、HOST TYPE 712と、状態713と、容量714がログボリュームグループで使用するLog用Volumeの論理記憶装置の数分登録される。

#### 【0046】

次に、図9を参照してVolume Pool構成情報156のテーブルの構成について説明する。

Volume Pool構成情報156は、ログボリュームグループで使用するログデータを格納する記憶装置の容量が足りなくなった場合に容量を拡張しやすいように、予めパスが定義されていないものやリザーブ状態の論理記憶装置を一括管理するためのテーブルである。

エントリ801には、記憶装置システム4の論理記憶装置番号が登録される。エントリ802には、エントリ801の番号で識別される論理記憶装置のHOST TYPEが登録される。エントリ803には、同じくその論理記憶装置の状態が登録され、エントリ804には、その論理記憶装置のリザーブ情報が登録される。エントリ805にはその論理記憶装置の容量が登録される。

#### 【0047】

図10は、ログボリュームグループ使用量管理情報155のテーブルの構成を示す。ログボリュームグループ使用量管理情報155は、ログボリュームグループごとに、ログデータを格納する記憶装置の使用量を監視するために用意される。エントリ901にはログボリュームグループ番号が登録され、エントリ902にはログボリュームグループ内の全空き容量が登録される。

#### 【0048】

エントリ903には、ログボリュームグループで使用するログ用ボリュームの論理記憶装置の番号が格納される。エントリ904にはエントリ901で識別される論理記憶装置



の容量情報が格納される。エントリ 905 にはライト I/O で使用された領域の容量が登録される。この領域の容量は、使用監視情報 908 のテーブルを参照して把握できる。すなわち使用監視情報 908 は、ログ用ボリュームの論理記憶装置 139 をビットマップで管理しており、ライト更新があった領域にビットを立てて管理する。この使用監視情報 908 のビットマップは記憶制御装置 131 の不揮発メモリ 137 内に格納されている。エントリ 906 にはライト更新がかかっていない残り容量情報が格納される。エントリ 907 には、ログボリュームグループの全記憶容量に対する空き容量の割合を示す情報が格納される。

#### 【0049】

図 11 は、LOG ID 管理情報 157 のテーブルの構成を示す。LOG ID 管理情報 157 は、記憶装置システム 3 から送信されるライト I/O による更新データを記憶装置システム 4 で受信して、I/O レベルのログを作成する時に、ログデータに付加される識別情報 (ID) を管理するテーブルである。すなわち、LOG ID 管理情報 157 のテーブルは、ログデータに付加されたログ ID カウンタを管理するために、ログボリュームグループ毎に最も古いログデータの ID 値を格納するエントリ 1001 と、そのデータが作成された時間情報を格納するエントリ 1002 と、そのデータが格納されたログボリュームのアドレス情報を格納するエントリ 1003 と、最も新しいログデータの ID 値を格納するエントリ 1011 と、そのデータが作成された時間情報を格納するエントリ 1012 と、そのデータが格納されたログボリュームのアドレス情報格納するエントリ 1013 から構成される。

#### 【0050】

コマンド処理プログラム 142 は、処理すべきコマンドがログデータの記憶される先の論理記憶装置 140 に対するライト I/O 又はマーク I/O である場合に、ロググループ毎に用意されるログ ID カウンタ 160 及びタイマ 161 を参照してカウンタ 160 の値 (ログデータ ID) 及びタイマ 161 の値 (時間) を付加してログ ID 管理情報を作成する。

#### 【0051】

次に、図 13 を参照して、ログデータのフォーマットについて説明する。ログデータは、記憶装置システム 4 が記憶装置システム 3 からのライト I/O もしくはマーク I/O 処理要求を処理する毎にキャッシュメモリ 136 上に作成され、その後記憶装置 140 に格納される。ログデータは、その先頭にマークフラグ 1201 が付加される。マークフラグ 1201 は、ホスト 1 と記憶装置システム 4 でシステムの状態を一意に識別する MARK 情報であるかを識別する識別子である。マークフラグ 1201 の後には、ログ ID (即ちログ ID カウンタの値) 1202 と、タイマ値 1203 と、データ長 1204 のそれぞれのエントリと、更にライトデータ又はマークデータが格納されるエントリ 1205 を備えて構成される。

#### 【0052】

マークデータの場合には、ログデータを格納するログボリュームのアドレス 1210 を格納するエントリと、ホスト 1 がマーク ID 1211 を格納するエントリと、マークを実行した時間 1212 を格納するエントリと、マークを要求したアプリケーションの識別情報 1213 が格納されるエントリより構成される。

#### 【0053】

一方、ライトデータの場合には、ライトデータを格納するログボリュームのアドレス 1220 を格納するエントリと、ライトデータ 1221 を格納するエントリより構成される。

尚、ログデータの作成については図 20 を参照して後述する。

#### 【0054】

次に、図 12 を参照して、MARK ID 管理情報 158 のテーブルの構成について説明する。

MARK ID 管理情報 158 は、MARK ID 情報を管理するテーブルであり、ログ

ボリュームグループ毎に最も古いマークシーケンス番号の値を格納するエントリ 1101 と、そのデータを作成した時間情報を格納するエントリ 1102 と、そのデータが格納されたログボリュームのアドレス情報格納するエントリ 1103 と、最も新しいマークデータの ID 値を格納するエントリ 1104 と、そのデータを作成した時間情報を格納するエントリ 1105 と、そのデータが格納されたログボリュームのアドレス情報格納するエントリ 1106 を備えて構成される。

#### 【0055】

コマンド処理プログラム 111 が、ホスト 1 が送信されるマークコマンドを受領すると、このマークコマンドを使用してマーク情報を記憶装置システム 3 に送信してきたものを受領する。マークコマンドであれば、図 20 に示すようにマーク情報をログデータ化するが、その際に、データ内のマーク ID 1211 を読み出してテーブルに格納して管理する。

#### 【0056】

次に、フローチャートを参照しながらそれぞれの処理動作について説明する。まず、全体の処理動作の概要について説明しておく。本実施形態による計算機システムでは、記憶装置システム 4 において、記憶装置システム 3 の正論理記憶装置 (M-VOL) のある時点のデータのバックアップデータ (以下「スナップショットデータ」) を有する副論理記憶装置 (R-VOL) を作成し、保持する。スナップショットデータが作成された時点以降に、ホスト 1 から受領するライト I/O 処理要求を実行する度に、記憶装置システム 3 で発生した更新データを記憶装置システム 3 から記憶装置システム 4 に送信する。記憶装置システム 4 では、ライト I/O 処理後のデータ (即ちライト更新データ) をログデータとして記憶装置 23 に記録する。

#### 【0057】

さらに、ホスト 1 は、自らが作成する任意の情報であるマークポイント情報 (以下「MP 情報」)、即ちマークを付加する時点の情報を記憶装置システム 3 に対して通知する。具体的には、ホスト 1 は、任意の時点、例えば記憶装置システム 3 と記憶装置システム 4 との間でのデータを一致させる処理 (シンク処理) を行う時に、MP 情報を記憶装置システム 4 のログデータに書込む。従って、MP 情報は、ホスト 1 と記憶装置システム 4 の両方で管理されることになる。これによって、ホスト 1 が指示する MP 情報及び記憶装置システム 4 内のログデータに格納された MP 情報を利用して、記憶装置システム 4 は、ホスト 1 が意図した時 (MP 情報作成時) から記憶装置システム 4 が保持していたデータを高速にリストアすることができる。

#### 【0058】

このような処理を実行するために、ホスト 1 は、あらかじめ、ログデータを取得する準備指示 (ログ取得開始準備指示)、及びログ取得開始指示を記憶装置システム 4 に送信する。これにより、記憶装置システム 4 は、ログデータの取得を開始し、ログモードとなる。その後、計算機システムは、上述した MP 情報の遣り取りを行う。

#### 【0059】

まず図 14 を参照して、ペア作成の処理動作について説明する。記憶装置システム 3 は、ホスト 1 からペアの作成を指示するコマンドを受領する (1301)。するとコマンド処理プログラム 111 は、ペア管理プログラム 114 を呼び出し、ペア作成要求コマンドの内容を渡す。ペア管理プログラム 114 は、コマンド情報を参照して、ペア作成指示対象のデバイス (記憶装置) のチェックを行う (1302)。ペア管理用の差分管理ビットマップ 123 のビットを全て「1」にして、コピーを行うためにコピーの進捗を管理するためのポインタを用意し、ペア管理差分ビットマップ 908 の先頭を設定する (1303、1304)。

#### 【0060】

ペア対象の記憶装置として、リモートの記憶装置システム 4 が指定されている場合には、リモートの記憶装置の検出を行い、通信路の確立を実行する (1305)。リモートの記憶装置の検出が出来ない場合 (1306) には、リトライ処理を実行する (1307)。

。そして最終的に、リモートの記憶装置の検出が出来なかった場合には、障害が起きたものとしてホスト1に応答する(1309)。尚、システムの設計上予め記憶装置システム3と記憶装置システム4の間の通信路が確立している場合には、この一連処理は必要なくなる。

#### 【0061】

リモートの記憶装置が検出される(1306)と、リモートの記憶装置との通信路の確立処理が行われる(1310)。その後、ペア管理情報の設定と通信路の確立が出来たら(1311)、記憶装置システム4に対してペア作成コマンドを送信する(1316)。そして記憶装置システム4からペア作成コマンドに対する応答を記憶装置システム3が受領したら(1317, 1318)、コピーマネージャプログラム113はコピー範囲を設定して(1319)、リモート側の記憶装置システム4に送信するデータを作成する。そしてリモートI/O処理プログラム112を実行してリモート側にそのデータを送信し、コピー進捗ポインタを進める(1320)。記憶装置システム4からデータ送信に対する完了報告が帰ってきたら、送信部分のビットマップ123を「0」クリアする(1321)。

そして、差分ビットマップ123が全て「0」かを判定して、もし全てがクリアされていれば(1322)、Pair管理テーブル118のペア状態604を“DUPLEX”に登録して(1323)、ホスト1にペア作成完了報告を返す(1324)。

#### 【0062】

尚、通信路の確立や記憶装置システム3と記憶装置システム4間でのコマンド、データのやり取りで、記憶装置システム4からの状態の確認要求、エラー報告等の不具合が起きた旨の情報が応答として返って来た場合や、コマンドに対する応答時間が長く、タイムオーバーした場合などはホスト1にエラー報告を返す。

#### 【0063】

次に、図15を参照して、ペア作成中のコマンド処理動作について説明する。ペア作成中にも記憶装置システム3はホスト1からデータの入出力及びその他システムの状態等を監視するコマンド等を随時受領している。この際のI/O処理要求を処理するコマンド処理プログラム111の動作を説明する。

#### 【0064】

コマンドを受領したコマンド処理プログラム111は、ライトI/O処理要求かを判断する(1401, 1402)。この判断の結果、ライトI/Oの場合には、ペア作成対象デバイス(記憶装置)か否かを確認する(1403)。もしペア作成対象のデバイスであれば、ライトI/O対象アドレスとペア差分管理ポインタのアドレスとの比較を行い(1404)、ライトI/Oを送信済みかどうか確認する(1405)。もし送信済みであれば、ライトI/Oデータをキャッシュメモリ106の所定のエリアを確保して格納する(1406)。そしてペア管理プログラム114を呼び出して、このライトI/Oデータを格納したキャッシュメモリ106のそのエリアの情報を渡す。ペア管理プログラム114は、優先的に当該ライトデータに対するリモート側への送信処理を行うために、ペア情報を用いて、リモートへのI/O処理コマンドとデータを作成して、リモートI/Oプログラムを用いてリモートの第2の記憶装置システム4にデータを送信する(1407)。記憶装置システム4からライト完了報告応答を受領したら(1408)、ホストにライト処理終了報告を返す(1409)。

#### 【0065】

一方、リモートペア対象外の場合(1403N)や、ペア作成のためのデータ送信が未だ送信されていない場合(1405N)には、キャッシュメモリ106のエリアを確保してデータを格納し(1411)、その後ホスト1にライト処理終了報告を返す(1409)。

また、ライトI/Oか否かの判定(1402)においてライトI/O処理以外の場合には、所定の処理を行い、ホスト1に処理終了報告を返す。例えばリード処理の場合には、リードI/Oの対象アドレスに対応するキャッシュメモリ又は記憶装置からデータを読み出

して(1421)、ホスト1にリードデータを送信した後、完了報告を返す(1422)。

#### 【0066】

次に、図16を参照して、マスタ側のログの取得開始処理について説明する。  
記憶装置システム3における記憶装置21と、第2の記憶装置システムにおける記憶装置22との間でペアを確立してDUPLEX状態になったら、ホスト1は、任意の契機で、被災のために誤ったデータがリモートに送信されても、リモート側のペアに即座に反映されないように、ペア中断コマンドとログデータ取得開始コマンドを送信する。

#### 【0067】

記憶装置システム3は、ホスト1から送信されるペア中断(サスペンド)コマンド及びログデータ取得開始コマンドを受領すると(1501)、その受領したコマンドを解析して、ペア管理プログラム114を呼び出して、ログボリュームグループ番号、ペア状態、デバイス(記憶装置)を確認し(1502)、リモートI/Oプログラム112を実行してリモートの記憶装置システム4にサスペンドコマンド及びログデータ取得開始コマンドを送信する(1511)。

#### 【0068】

続いて記憶装置システム3は、記憶装置システム4から送信されてくるログ取得準備完了の応答の受信を待つ。そして準備完了の応答を受領したら(1503)、記憶装置システム4側のペア管理プログラム145は、指示されたペアに関する記憶装置番号を基にペア構成情報118のペア番号を割り出し、ログボリュームグループ内のこのペアの状態をサスペンドに設定する(1504)。その後、ログ取得コマンドに対する完了報告をホスト1に返す(1505)。

一方、ログ取得準備完了の応答ではなく(1503)、不具合な応答を受領した場合(1521)には、ホスト1にエラー報告をして処理を終わる(1522)。

#### 【0069】

次に、図17を参照して、リモート側のLOG取得開始処理について説明する。  
リモート側の記憶装置システム4は、記憶装置システム3からペアサスペンドコマンド及びログデータ取得開始コマンドを受領すると(1601)、受領したコマンドを解析して、ペア管理プログラム145を呼び出してログボリュームグループ番号、ペア状態およびデバイスを確認する(1602)。

続いて、ボリューム構成情報テーブル152を参照してペア確認を行い、関連するペア管理情報テーブル153をサスペンドにして(1603)、ログボリューム用記憶装置(リザーブ)が用意されているかを確認する。この時、ログボリューム用リザーブが用意されていなければ、論理記憶装置140を割り当てる。その後ログボリュームグループ用のログシーケンスカウンタを初期化し(1604)、ログボリュームグループ情報フラグをONにして、完了報告を記憶装置システム3に応答する(1605)。

#### 【0070】

次に、図18を参照して、MARKコマンドの処理動作について説明する。  
ホスト1と記憶装置システム4の両方で同一のマーク情報を保持するために、ホスト1から発行されたマークコマンドは、記憶装置システム3を通して記憶装置システム4に送信される。

#### 【0071】

図18Aに示すように、ホスト1側では、ホスト1のAPやDBMS(データベース管理システム)は、ファイルのセーブ時、ファイルのクローズ時、APのセーブ時、APのプロセスコミット時などの後にマーク処理プログラム243を呼び出してマークコマンドを発行する(1701)。マークコマンドを発行する際にマーク処理プログラム243はマークカウンタとタイマを参照し、マークIDとタイマ値を挿入して、マークデータを作成する。そして記憶装置システム3からReady応答を受領すると(1702)、記憶装置システム3へマークデータを送信する(1703)。その後一連の処理が終了して、記憶装置システム3から完了報告を受領すると、マークコマンドの処理を終える(170

4)。

【0072】

次に、図18Bを参照して記憶装置システム3におけるMARKコマンドの処理について説明する。

記憶装置システム3のコマンド処理プログラム111は、ホスト1から送信されるマークコマンドを受領して(1705)、そのコマンドを解析する(1706)。解析の結果、もしそのコマンドがマークコマンドならば、ホスト1へReady応答を返す(1707)。

【0073】

記憶装置システム3は、ホスト1からコマンドに続いて送信されるマークデータ又はライトデータを受領する(1708)。そして受領したコマンドに応じた処理を行う(1709)。もし、マークコマンドの場合には、リモートの記憶装置システム4へマークコマンドを送信する(1710)。そして記憶装置システム4からReady応答を受領したら(1711)、記憶装置システム4へマークデータを送信する(1712)。マークデータを送信した後、記憶装置システム4ではマークデータの処理が行われる。そして、処理が終わると、記憶装置システム3へ完了報告を送信する。記憶装置システム3は、リモートの記憶装置システム4から完了報告を受領すると(1713)、記憶装置システム3は、ホスト1にマークコマンド処理完了報告を返す(1714)。

【0074】

ところで、先のステップ1706の判定において、受領したコマンドがリードコマンドである場合には(1715)、まずキャッシュのヒット/ミスヒット、即ちキャッシュメモリ106に目的とするアドレスの有無を判定する(1716)。そして、キャッシュメモリ又は記憶装置21から目的とするデータを読み出して(1718)、ホスト1へ送信する(1718)。一方、リードコマンドで無ければ所定の処理を行う(1719)。また、上記ステップ1709の判定で、ライトコマンドならば、キャッシュのヒット/ミスヒットを判定し(1720)、ライトデータを対応する記憶装置109及びキャッシュメモリ106に書き込み、非同期処理のキューに登録する(1721)。そしてリモート側へ送信すべき対象のデータか否かを判定し(1722)、送信すべきデータならばステップ1710へ移り、以後の処理を行う。リモート送信データで無ければ、完了報告をホスト1へ送信して終了する。

【0075】

次に、図19を参照して、リモートサイトの記憶装置システム4におけるログ取得中のコマンド処理について説明する。

記憶装置システム4は、記憶装置システム3からコマンドを受領すると(1801)、マークコマンドかライトコマンドか、それともそれ以外かを判定する(1802)。マークコマンドを受領した場合には、マークコマンドに対するReady応答を記憶装置システム3に返す(1803)。

【0076】

マスタ側の記憶装置システム3では、リモート側からReadyコマンドを受領すると、リモートにマークデータを送信する(1712)。リモート側の記憶装置システム4では、マークコマンドの受領に引き続いてマークデータを受領すると(1804)、ログ取得対象デバイス(論理記憶装置)か否かを判定する(1805)。この判定は、リモート側のボリューム構成情報152、ペア管理情報153、ログボリュームグループ構成情報154のテーブルを順に参照することにより行われる。この判定の結果、対象のデバイスがログデータ取得対象のデバイスであるならば、ログデータ作成処理を行う(1806)。尚、ログデータ作成処理については図20を参照して後述する。ログデータの作成が完了すると、マスタ側の記憶装置システム3へ完了報告を返す(1807)。

【0077】

ところで、ステップ1802の判定の結果、リードコマンドを受領した場合には、リードコマンドの処理(1808～1812)を行い、ライトコマンドのときはライトコマン

ドの処理(1813~1814)の行う。これらの処理は、前述した図18に示す動作と同様であるので、説明を省略する。

#### 【0078】

次に、図20を参照して、記憶装置システム4におけるログデータ作成処理について説明する。

ログデータ作成処理では、まずログID管理情報テーブル157を参照しLOG Volumeのアドレスを設定し(1901)、ログIDカウンタ160及びタイマ161を参照してログIDとタイマ値を設定する(1902)。そしてログデータ格納用にキャッシュメモリ136内に領域を確保する(1903)。その後、図13(A)に示すようにマークデータ(マークID1211及びマーク時間1212等を含む)に、ログ情報(ログID1202及びログ時間1203等を含む)を付加したログデータをキャッシュメモリ136の領域に格納する(1904)。後で非同期処理にこのログデータ作成処理を追加する(1905)。即ち、非同期記憶装置I/O処理プログラムの実行によりキャッシュメモリ136の当該領域に格納されたログデータは、記憶装置23内の指定された論理記憶装置140に格納される。その後記憶装置システム3に完了報告を返して終了する。

#### 【0079】

次に、図21を参照して、LOG Volume容量オーバー処理について説明する。ログボリュームグループ管理プログラム146は、ロググループ毎にログボリュームの使用量を監視する。ライトI/OもしくはマークI/Oによる更新が生じた場合には、図10に示すログボリューム使用量管理情報155に反映される。更にこの際に反映したログボリューム使用量管理情報155を参照して(2002)、ログボリュームの使用量が規定値以上の量を使用していないかを確認する(2003)。尚、この規定値は、ホスト1又は管理装置8、9もしくは計算機システム管理装置13を用いて設定されている。上記判定の結果、規定値以上の使用量であれば、通信路6を介して記憶装置システム3に対してログボリュームの容量不足の旨のAlertを出す(2004)。

#### 【0080】

このAlertを受けた記憶装置システム3は、このAlertをホスト1に返す。または、第2の記憶装置管理装置4へAlertを出し、第2の管理装置9から計算機システム管理装置13やホスト1にAlert情報が伝達されるようにする。この場合、システム管理者は、ログボリュームを反映するように記憶装置システム4に指示をする。ホスト1から指示を出す場合は、ホスト1のマークID管理情報を参照して、適当なマークIDを指示してログボリューム内のログデータをR-VOLに反映して、ボリューム容量不足を回避するか、容量追加を指示する。それを受領した記憶装置システム4は、Volume Pool構成情報156に登録されている空きVolumeを割り当てることによりボリューム容量不足を回避する。

#### 【0081】

次に、図22を参照して、Take Over処理について説明する。

この処理はホスト1、2と連携して行われる処理である。

まず、ホスト2のクラスタソフトウェアは、ホスト1のクラスタソフトウェアとの間で通信路10を介してハートビートの監視を行う(2101)。具体的には、所定の時間間隔に、決まったあるデータをホスト1とホスト2のクラスタソフトウェアは通信しあい、ある所定の時間内での応答を期待する。

この監視において、ハートビートに対する応答が所定回数無い場合には、相手方ホストの障害かネットワーク通信路の障害が想定されるので、ハートビート切れと判断して(2102)、交替系の通信路を使用して更にハートビートを行う(2103)。尚、交替系の通信路は、通常その機能を満たすためには、正常系の通信路と物理的に分かれていたほうが望ましい。

#### 【0082】

交替系の通信路でも、応答がない場合には(2104)、システム管理者に障害が発生した可能性があるためTake Overを行うための問い合わせメッセージを出す(2

104)。システム管理者は、ホスト1があるサイトで何らかの災害が発生して障害が起きた場合で、ホスト1がまだ計算機処理を行える状態であった場合は第1のクラスタソフトを用いてTAKE OVER処理を実行する。具体的には、記憶装置システム3のリモートペアを全てサスペンドにするコマンドを記憶装置システム3に送信すると共に、第2のホストに計算機システム4の処理を交代する。この際に処理中のアプリケーションは中断され、リモート間のコピー処理は中断する(2106)。

#### 【0083】

ホスト2においても、記憶装置システム4に対して、サスペンド処理を行うコマンドを送信する。もしくはホスト1との間でのハートビートが、一定時間途絶えたら、ホスト2のシステム管理者はホスト1のシステム管理者に問い合わせ、ホスト1側が被災していないかを確認し、被災していることが判明した場合や、既にマスコミ媒体によって被災の情報が知れている場合には、ホスト2から記憶装置システム4にペアサスペンドコマンドを投げ、ペアを中断する。そしてペア情報はサスペンドに設定される(2107)。しかし、被災中のためホストの情報処理のためのメモリの状態が異常になるなどで誤ったデータを書き込む処理を行う可能性がある。この場合も記憶装置システム3は、記憶装置システム4へデータを投げる。以後、ホスト1又は2は記憶装置システム4を用いてシステムのリカバリを行う。

#### 【0084】

次に、図23を参照して、リカバリ処理について説明する。  
このリカバリ処理とは、記憶装置システム4がホスト2からリカバリリストア指示を受領して行うデータ回復のための処理をいう。尚、以下の処理は、記憶制御装置131が、コピーマネージャプログラム144とログID管理プログラムを実行することで行われる(2201)。

ホスト1のリカバリ処理を行うホスト2にとって、論理的不整合等の障害が起きる可能性があるデータを、記憶装置システム4が記憶装置システム3から受け取っている可能性がある。そこで、ホスト2からリカバリ処理を行う場合に、記憶装置システム3の記憶装置M-VOLとペアを構成している記憶装置システム4のR-VOL及びログ論理記憶装置140に格納されたデータを使用し、ホスト2から記憶装置システム4に対して、リストアすべき論理ボリュームの識別子とR-VOLとマークポイントを用いたリストア処理要求を送信する(2202)。

#### 【0085】

ホスト2は、リストア処理要求を受領する論理記憶装置を確認する(2203)。これは、記憶装置システム4に対して、マークID管理情報を取得するためのコマンドを送信し、マークID管理情報を取得することにより行う。取得されたマークID管理情報は、例えばホスト2の表示器、又は管理装置9の表示器に表示され、システム管理者によりリストア先の論理記憶装置が確認される(2204)。即ち、システム管理者は、表示されたマークID管理情報を参照し、回復ポイントを決める。そしてこれに対応するマークIDを選択し、回復コマンドの内容に含め送信するようにホスト2を用いて指示する。  
ここで、ホスト2のシステム管理者は、障害発生直前のマークIDではなく、マークID情報のリストから、任意のマークIDを選択することができる。これにより、システムのユーザは、任意のマークIDを選択することで、選択されたマークIDが作成された時点に記憶装置システム4のR-VOLの状態を復元することが出来る。

#### 【0086】

ホスト2は、ステップ2201で選択したマークIDまでのログデータのリストア処理要求を記憶装置システム4に発行する(2205)。リカバリ処理となるリストア処理要求には、ログデータを反映する処理の対象となるR-VOLの識別子(例えばWWNとLUN等)、R-VOLの属するログボリュームグループを指定する識別子、選択されたマークIDの情報等が含まれる。ホスト2より発行されたリストア処理要求を受領した記憶装置システム4の記憶制御装置131は、システム構成情報管理プログラム163を実行して、リストア処理要求に含まれるログボリュームの識別子とログボリュームグループ情

報を比較参照し、指定されたログボリュームがR-VOLに対する正しいログボリュームであるかを確認する。(2206)。

【0087】

更に、記憶制御装置131は、リストア処理要求の内容から、R-VOLにログデータの反映処理を行うのか、或いは異なった未使用の記憶装置109にリストア処理を行うのかを確認する。尚、記憶装置109の障害により処理続行が出来ない場合はその旨をホスト2に通知し、処理を中止する。

記憶装置21以外の他の記憶装置へデータをリストアする場合には、記憶制御装置131は、キャッシュメモリ136にデータ格納領域を確保する。その後、記憶制御装置131は、コピーマネージャプログラム144を実行して、R-VOLに対応するログボリュームの先頭から、順次ログデータをキャッシュメモリ136に確保された領域に読み出す(2207)。その際、読み出されたログデータにマーク情報が含まれるかどうかを確認する(2208)。即ちログデータのマークフラグがONになっているかどうかを確認する。

【0088】

読み出されたログデータが、マーク情報を含むログデータである場合、記憶制御装置131は、更に読み出されたログデータに含まれるマークIDがホスト2から指定されたマークIDかどうかを確認する(2209)。ログデータに含まれるマークIDがホスト2から指定されたマークIDでない場合、又はログデータにマークIDが格納されていない場合(マークフラグ(MP)がONになっていない場合)には、記憶制御装置131は、読み出されたログデータに含まれるライトデータを、R-VOL又はその他の記憶装置の対応するアドレスに書き込むように送信する。一方、マークIDに対応するログデータである場合には、ライトデータが存在しないので、データの書き込みは行われない(2212)。

その後、記憶制御装置131は、ステップ2207に戻り、次のログデータを読み出し処理する。以下、記憶制御装置131は、2207~2211の処理を繰り返すことで、指示されたマークIDまでのログデータを指定された記憶装置のアドレスにリストアする。

【0089】

上記ステップ2209で、マークIDが指定されたマークIDと一致した場合には、記憶制御装置131は、リストアすべきデータをすべてR-VOLや他の記憶装置109に書き込んだと判断して、リストア処理の終了をホスト2に通知する。

ホスト2は、記憶装置システム4から終了報告を受領したら、ホスト2が指定したマークID時点までのデータが回復されたと判断して、他の処理を継続する。記憶装置システム4は、ホスト2からリカバリ処理の完了応答を受領すると(2212)、処理を終了する。

【0090】

以上、本発明の一実施例について説明したが、本発明は上記以外にも種々変形して実施し得る。

例えば上記実施例では、ログデータをリモートサイトの記憶装置システム4内に記憶するように構成している。しかし変形例では、リモートサイト側だけでなく、マスタサイトの記憶装置システム3内の記憶装置にもログデータを記憶するようにしてもよい。この場合には、図2に示した記憶装置システム4側に備えられたプログラムや情報は、マスタサイトの記憶装置システム3にも具備することは明らかなる。

【0091】

また、リストア先となる記憶装置は、必ずしもマスタサイトの記憶装置システム3内の記憶装置21とは限らない。マスタサイトが被災状態にあるときには、リモートサイトの記憶装置システム4内にある記憶装置139内のボリューム139でもよい。或いはネットワークに接続された他の記憶装置システム内の記憶装置でもよい。

【0092】

通常、リストア先の記憶装置にログデータをリストアした後は、マスタサイトの記憶装



置が保持していたデータが復元されたものとして扱われるので、その時点までのログID管理情報157、マークID管理情報158、及びログデータ等は、捨てられてよい。しかしながら、データの改ざん防止のため、或いは一定期間履歴として残しておくべき規則などがある場合には、上記管理情報及びログデータはそのまま保持し続けてもよい。

【図面の簡単な説明】

【0093】

- 【図1】一実施形態による計算機システムの構成を示すブロック図。
- 【図2】記憶装置システム3及び4の構成を示すブロック図。
- 【図3】ホスト1の構成を示すブロック図。
- 【図4】記憶装置管理情報115、150のテーブルの構成を示す図。
- 【図5】記憶装置構成情報116、151のテーブルの構成を示す図。
- 【図6】Volume構成情報117、152のテーブルの構成を示す図。
- 【図7】ペア管理情報118、153のテーブルの構成を示す図。
- 【図8】ログボリュームグループ情報110のテーブルの構成を示す図。
- 【図9】Volume Pool構成情報156のテーブルの構成を示す図。
- 【図10】Log Volume使用量管理情報155のテーブルの構成を示す図。
- 【図11】LOG ID管理情報157のテーブルの構成を示す図。
- 【図12】MARK ID管理情報158のテーブルの構成を示す図。
- 【図13】ログデータのフォーマットを示す図。
- 【図14】ペア作成の処理動作の説明に供するフローチャートを示す図。
- 【図15】ペア作成中のコマンド処理動作の説明に供するフローチャートを示す図。
- 【図16】マスタ側のログ取得開始処理動作のフローチャートを示す図。
- 【図17】リモート側のLOG取得開始処理動作のフローチャートを示す図。
- 【図18A】MARKコマンドの処理動作のフローチャート（ホスト側）を示す図。
- 【図18B】MARKコマンドの処理動作のフローチャート（記憶装置システム3側）を示す図。
- 【図19】リモートサイトにおけるログ取得中のコマンド処理のフローチャートを示す図。
- 【図20】リモートサイトにおけるログデータの作成処理動作のフローチャートを示す図。
- 【図21】LOG Volume容量オーバー処理動作のフローチャートを示す図。
- 【図22】Take Over処理のフローチャートを示す図。
- 【図23】リカバリ処理のフローチャートを示す図。

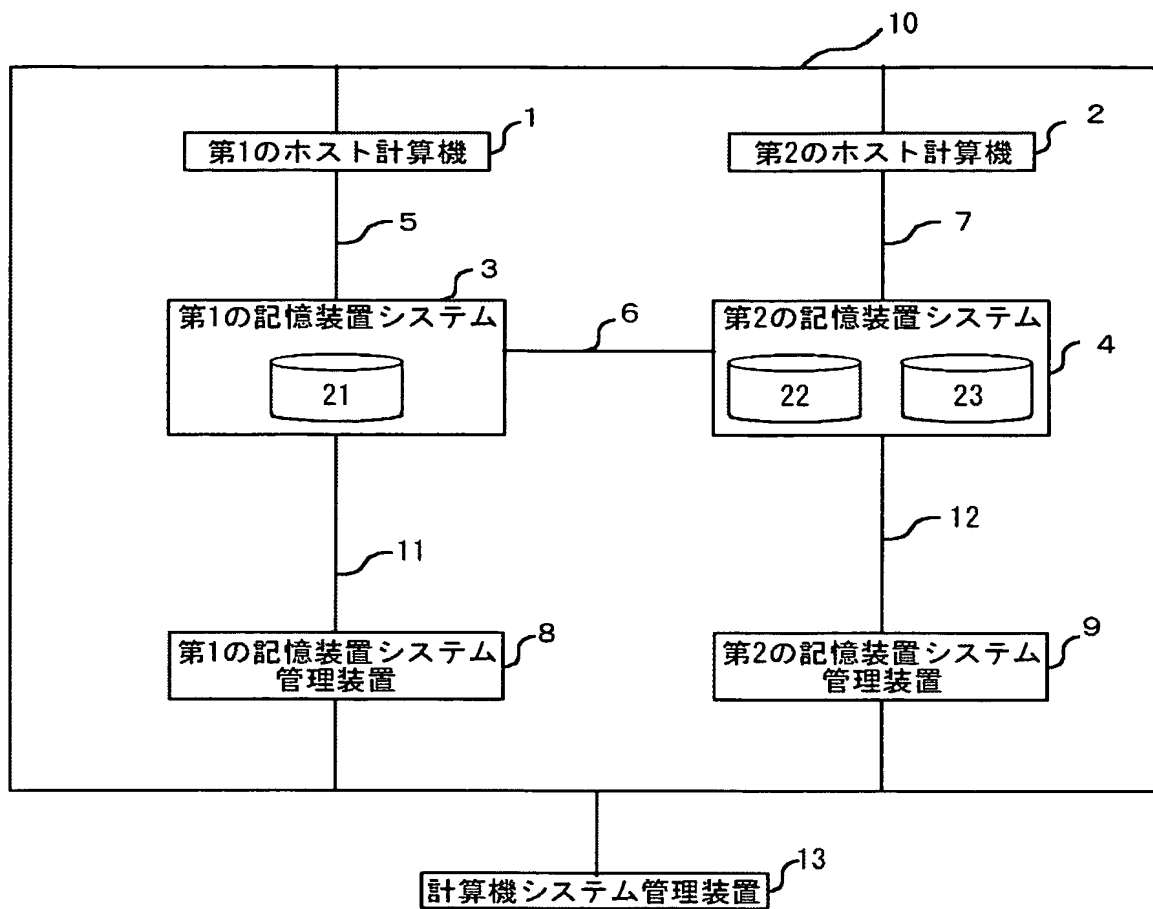
【符号の説明】

【0094】

1：ホスト、 2：ホスト、 3：第1の記憶装置システム、 4：第2の記憶装置システム、 8、9：記憶装置システム管理装置、 13：計算機システム管理装置、 5、6、7、10、11、12：通信路、 21、22、23：記憶装置、 109、139、140：記憶装置、

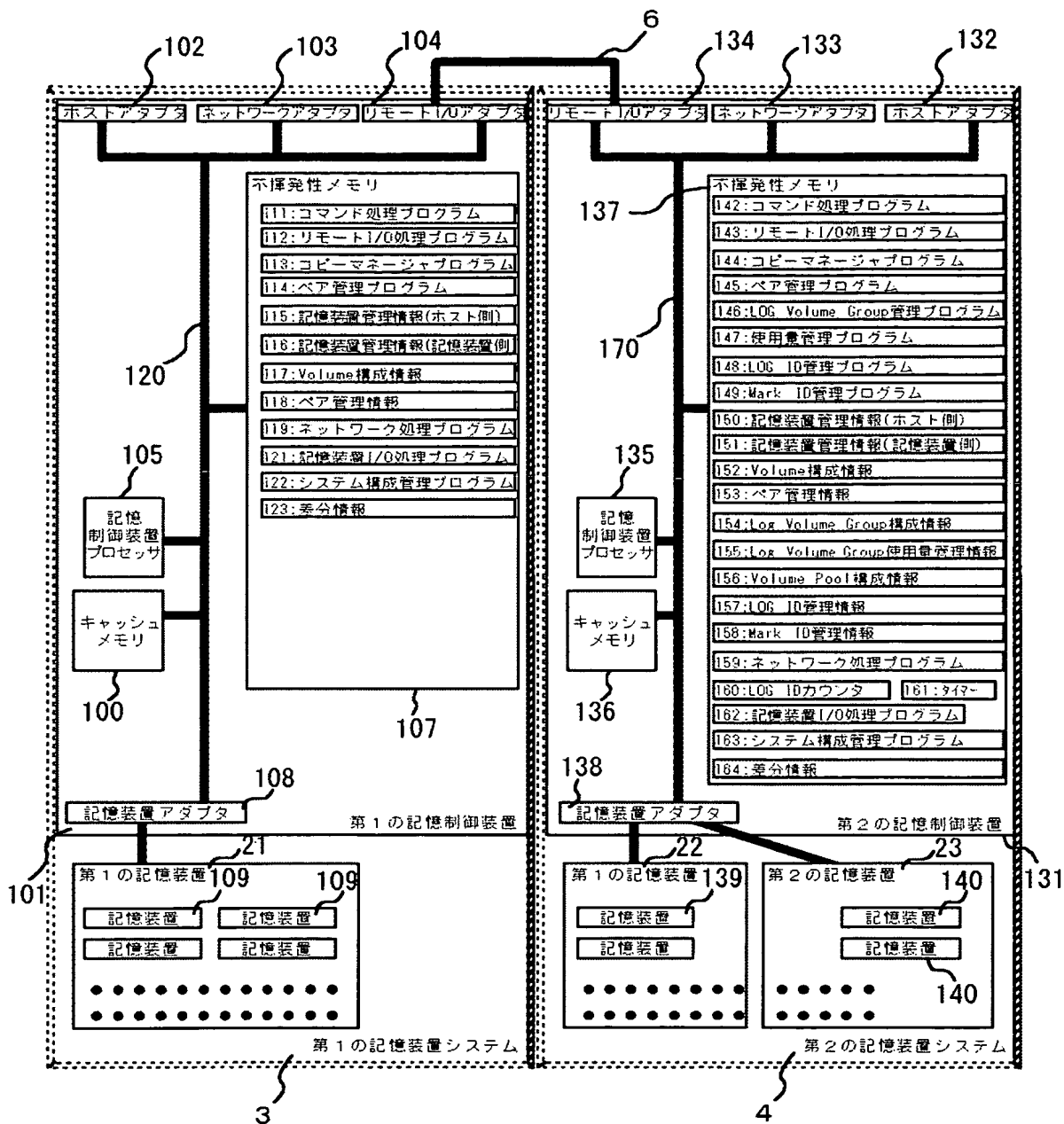
【書類名】図面  
【図 1】

図 1



【図 2】

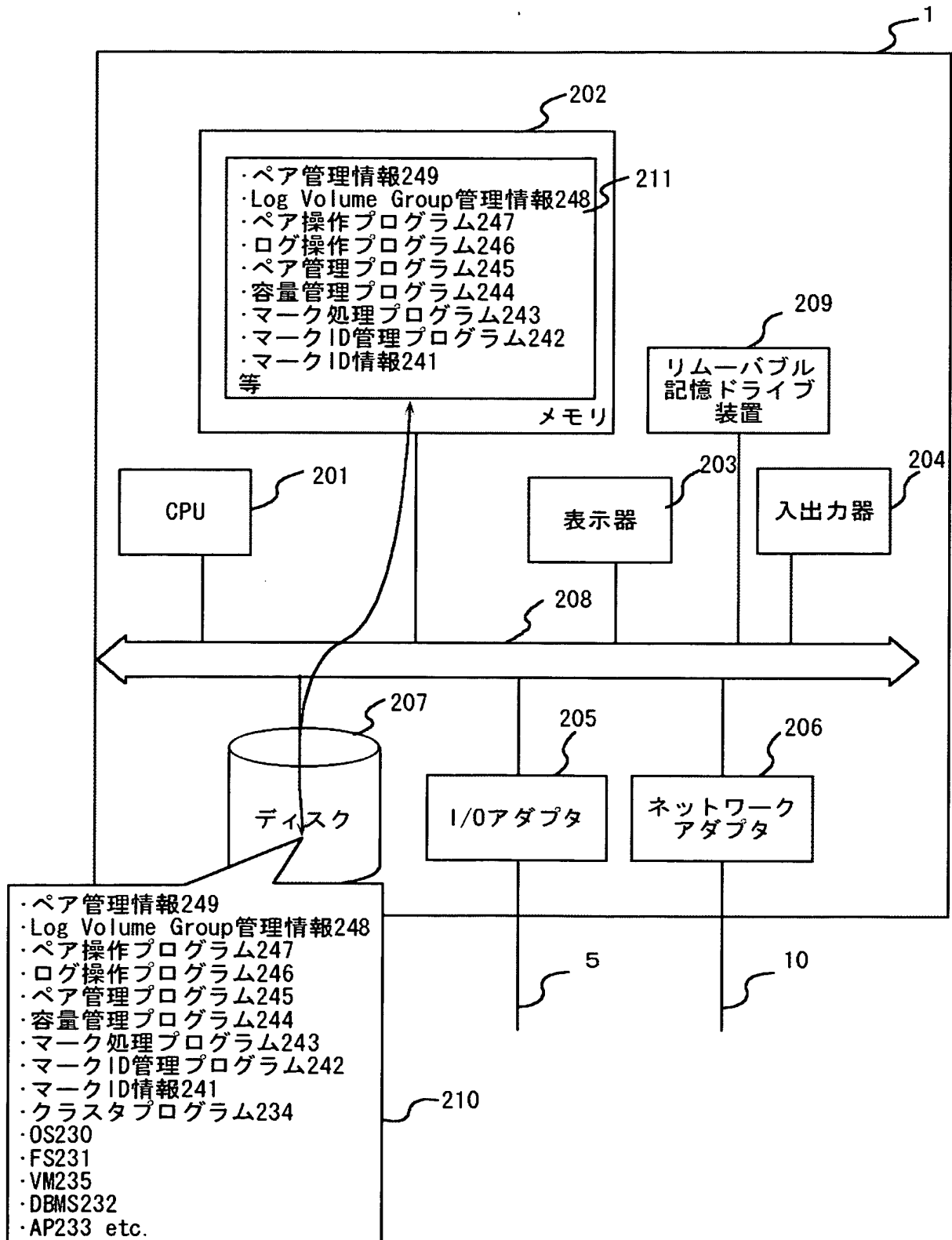
図 2



【図 3】

図 3

ホスト計算機ブロック図



【図 4】

図 4  
記憶装置管理情報 115

ホスト計算機の識別子：XXXXXXXXXX 301

ホスト計算機が認識する記憶装置のアドレス		記憶装置システム内部での論理的なアドレス	
記憶装置の識別子 303	記憶装置内アドレス 304	記憶装置システム内論理 記憶装置番号 305	記憶装置システム内論理 記憶装置番号内アドレス 306
aaaaaaaaaaaaa	000000000000	00	000000000000
aaaaaaaaaaaaaab	000000000000	0B	000000000000
aaaaaaaaaaaaaab	100000000000	0C	000000000000
.....	.....	.....	.....

【図 5】

図 5  
記憶装置管理情報 116

記憶装置システム内部での論理的なアドレス		RAID Groupに関するアドレス		ディスクに関するアドレス	
記憶装置システム内論理番号	記憶装置システム内論理アドレス	RAID Group番号	RAID Group内仮想アドレス	ディスク番号	ディスクアドレス
404	405	406	407	408	409
00	000000000000	00	000000000000	00	000000000000
00	000002000000	00	000002000000	01	000002000000
.....	.....	.....	.....	.....	.....

【図 6】

図 6

Volume構成情報 117.152

501

記憶装置 システム内 論理記憶 装置番号	Host Type 502	パス定義 503	状態 504	リザーブ情 報 505	Pair番号 506	ログ ボリューム グループ番号 507	容量 508

【図 7】

図 7

Pair 管理情報 118. 153

Pair 番号 601	記憶装置 システム 3 内 論理記憶 装置番号 (正) 602	記憶装置 システム 4 内 論理記憶 装置番号 (副) 603	状態 604

【図 8】

図 8

Log Volume Group 構成情報 154

フラグ 701			
Log Volume Group 番号: xxx 702			
論理記憶装置数: yy 703			
Log Volume Group 内論理記憶装置容量総和: zzz 704			
記憶装置シス テム内論理記 憶装置番号	Host Type 706	状態 707	容量 708
705 ...	...	...	...
Log Volume Group で使用する Log 用 Volume の論理記憶 装置識別子数 709			
Log Volume Group で使用する Log 用 Volume の論理記憶 装置の容量総和 710			
記憶装置シス テム内論理記 憶装置番号	Host Type 712	状態 713	容量 714
711 ...	...	...	...



【図 9】

図 9

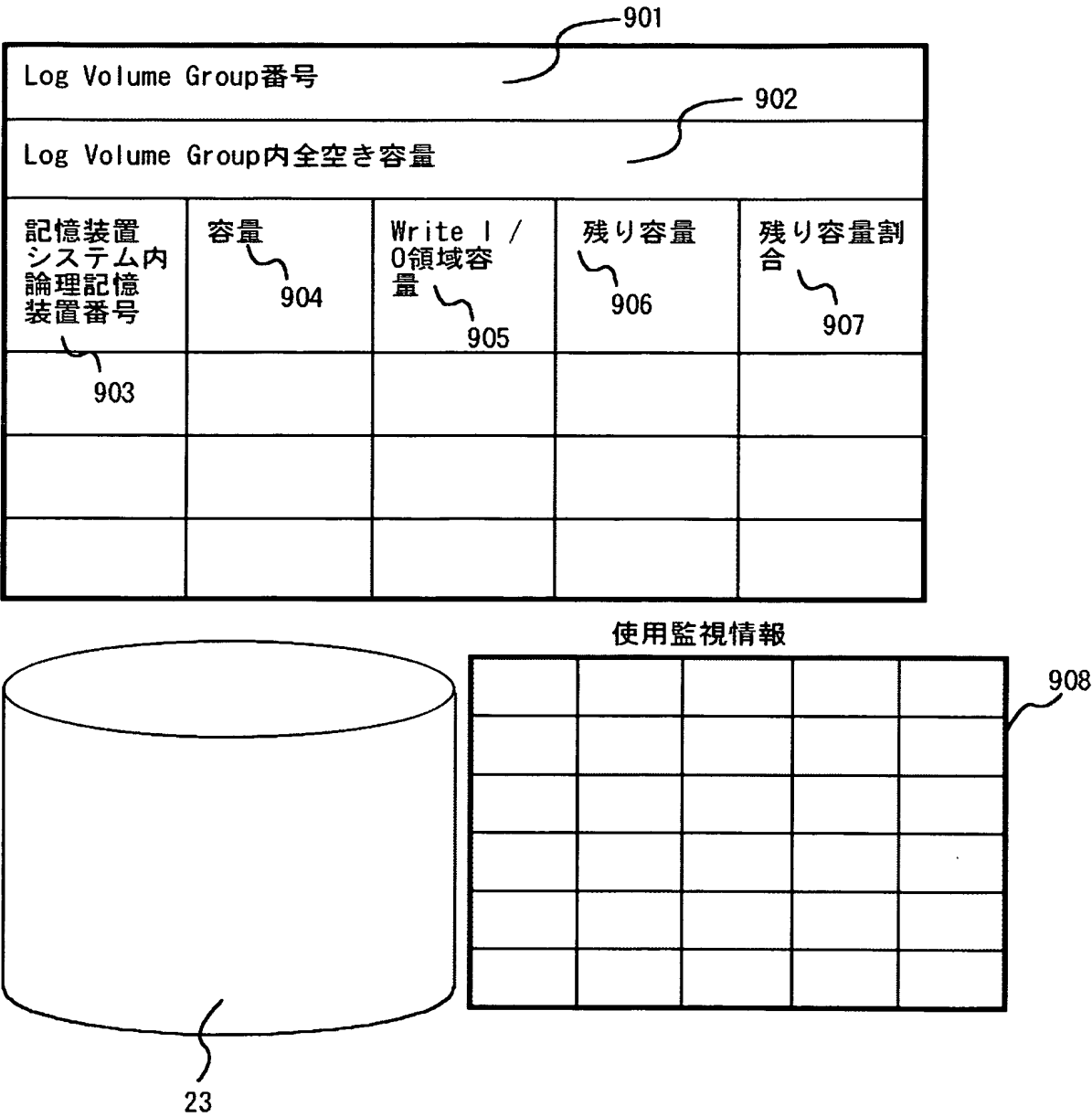
## Volume Pool構成情報

記憶装置システム内論理記憶装置番号	Host Type 802	状態 803	リザーブ情報 804	容量 805
801				

【図 1 0】

図 1 0

Log Volume Group使用量管理情報 155



【図 11】

図 11

## LOG ID 管理情報 157

最も古いLOG Data ID	1001
最も古いLOG Data の時間	1002
最も古いLOG Data のアドレス	1003
最新のLOG Data ID	1011
最新のLOG Data の時間	1012
最新のLOG Data のアドレス	1013

【図 12】

図 12

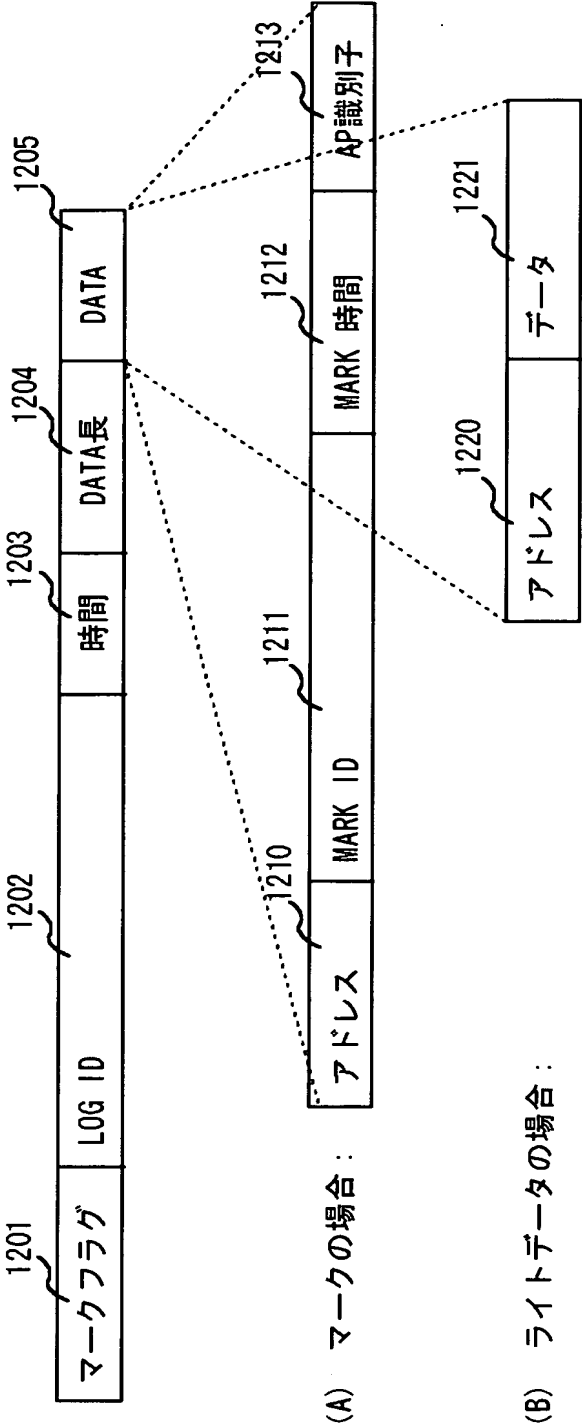
## MARK ID 管理情報 158

最も古いMARK ID	1101
最も古いMARK の時間	1102
最も古いMARK のアドレス	1103
最新のMARK ID	1111
最新のMARK の時間	1112
最新のMARK のアドレス	1113

【図 13】

図 13

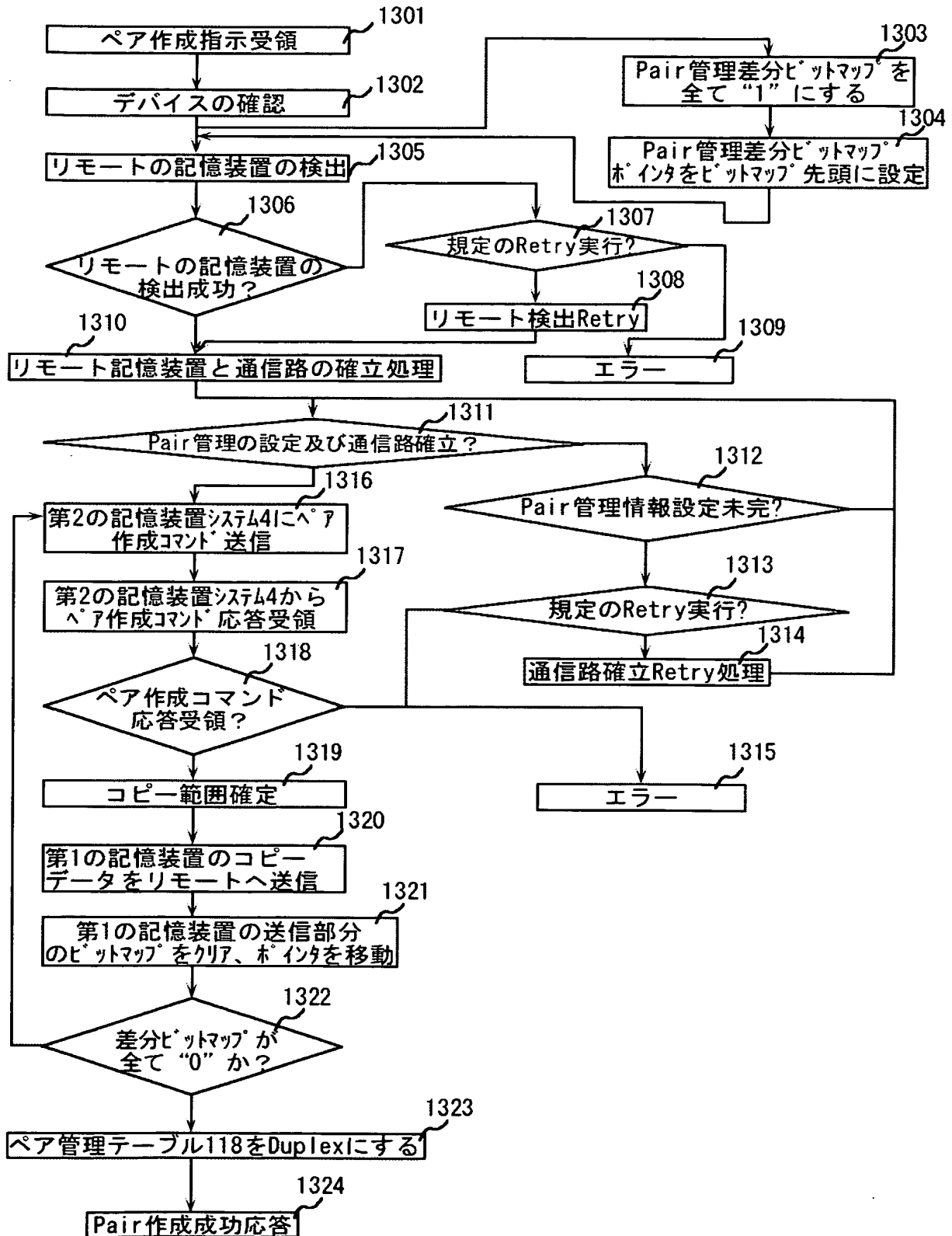
ログデータフォーマット



【図 14】

図 14

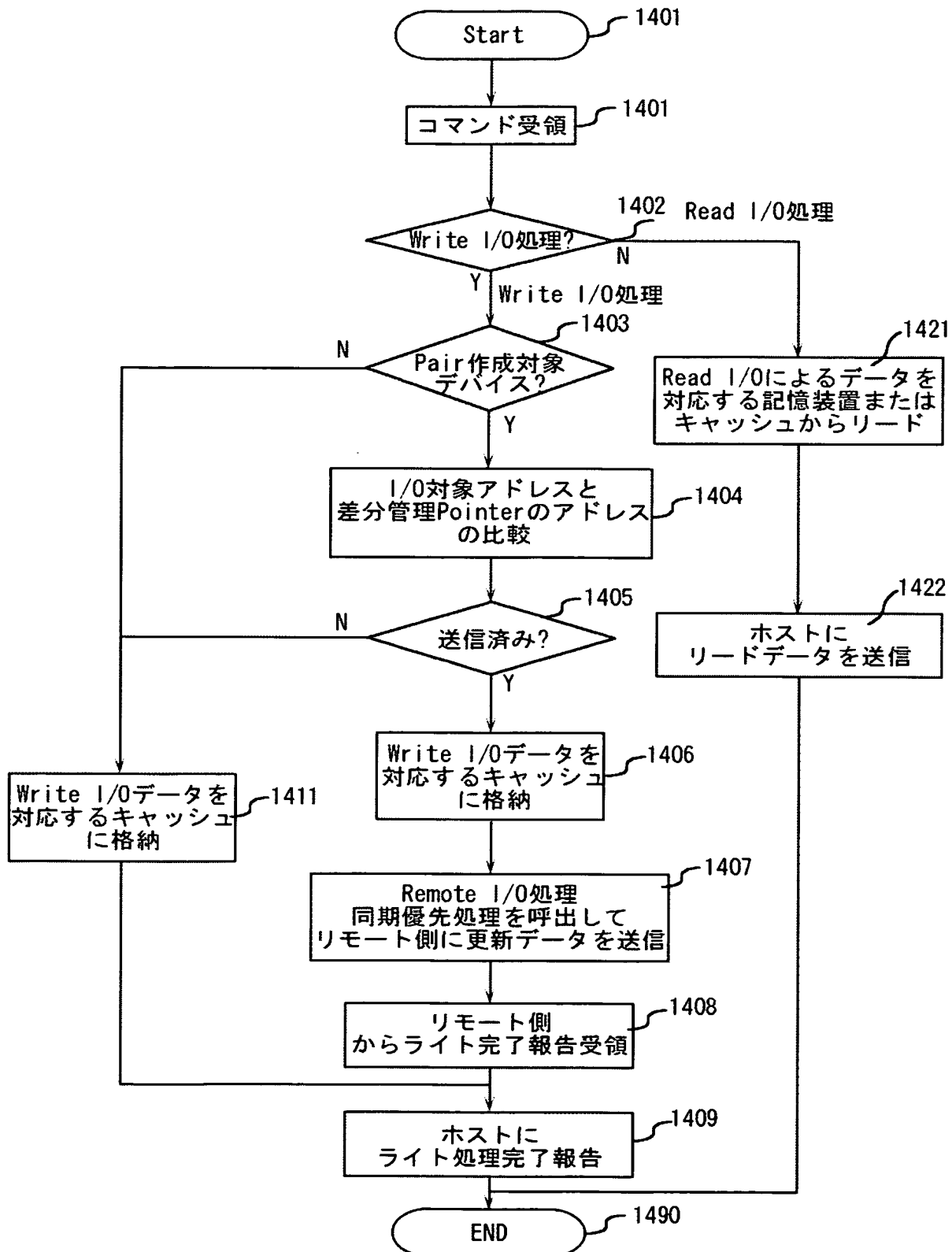
## Pair作成処理



【図 15】

図 15

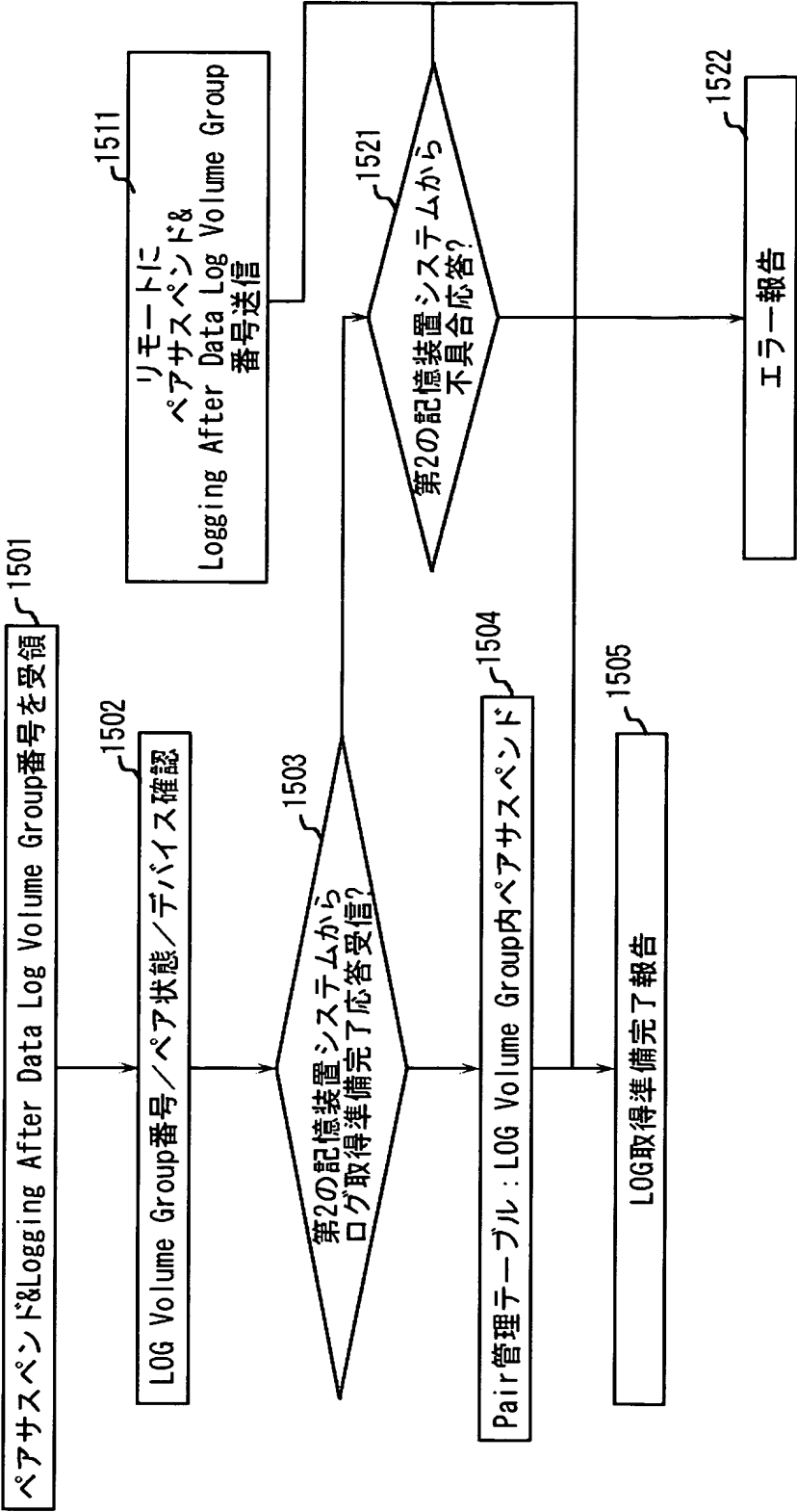
コマンド処理 (Pair作成中)



【図 16】

図 16

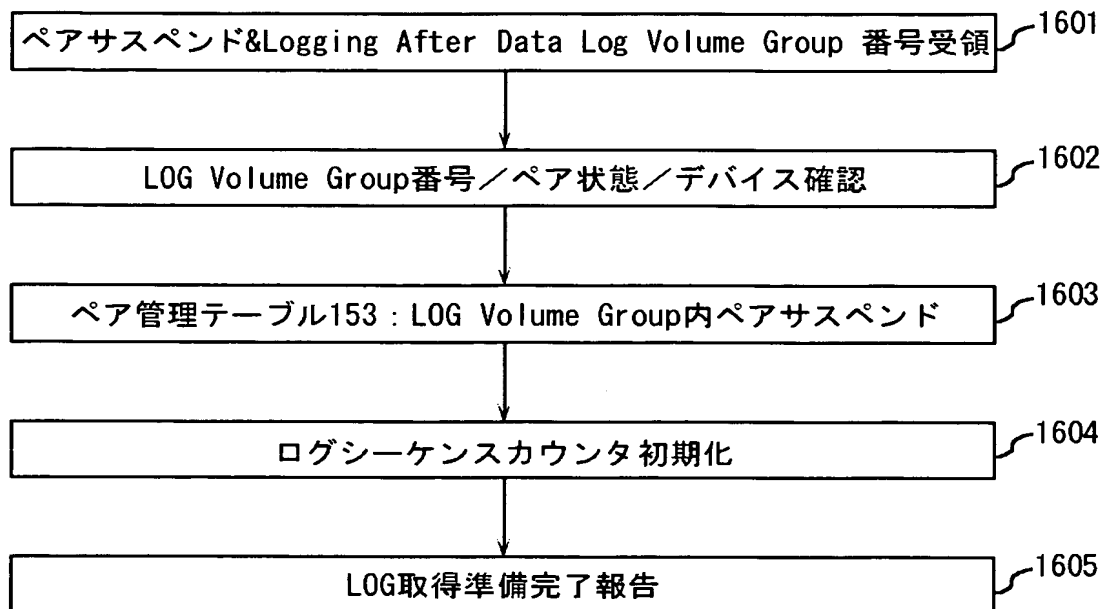
Log取得開始処理 (マスター側)



【図 17】

図 17

## Log取得開始処理（リモート側）



【図 18 A】

図 18 A

## MARKコマンド処理（ホスト側）

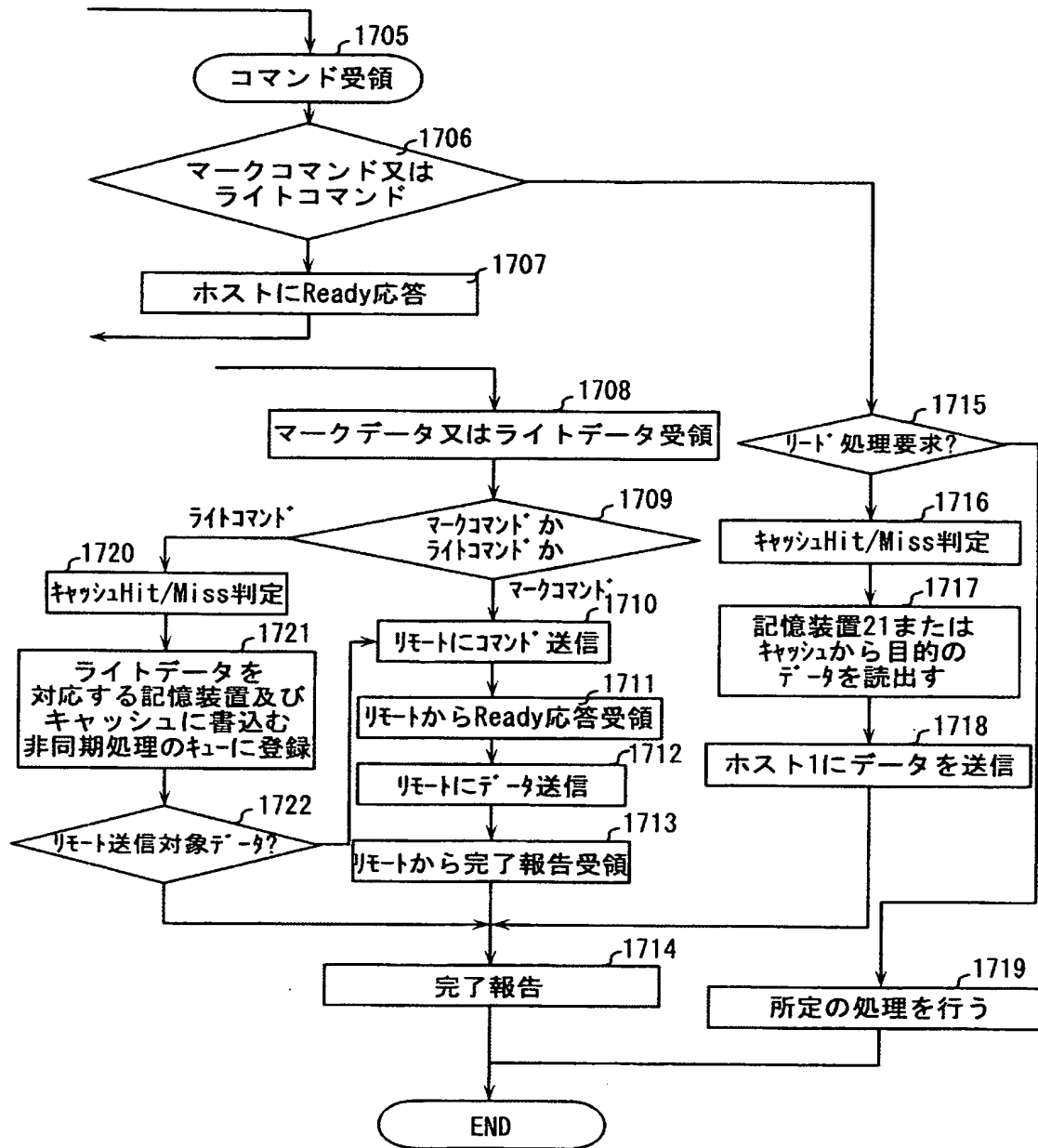




【図18B】

図 18B

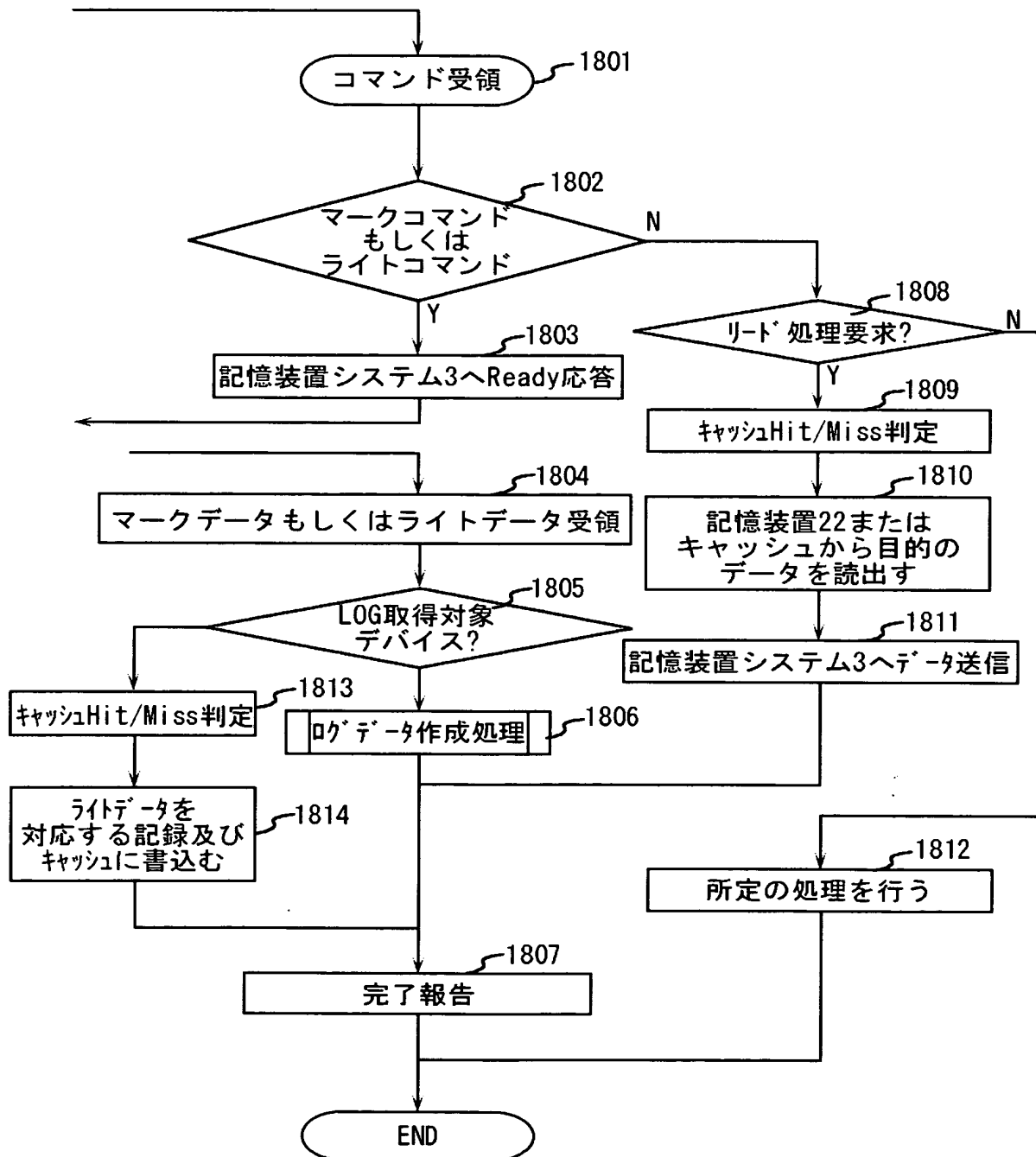
## MARKコマンド処理（マスター側）



【図 19】

図 19

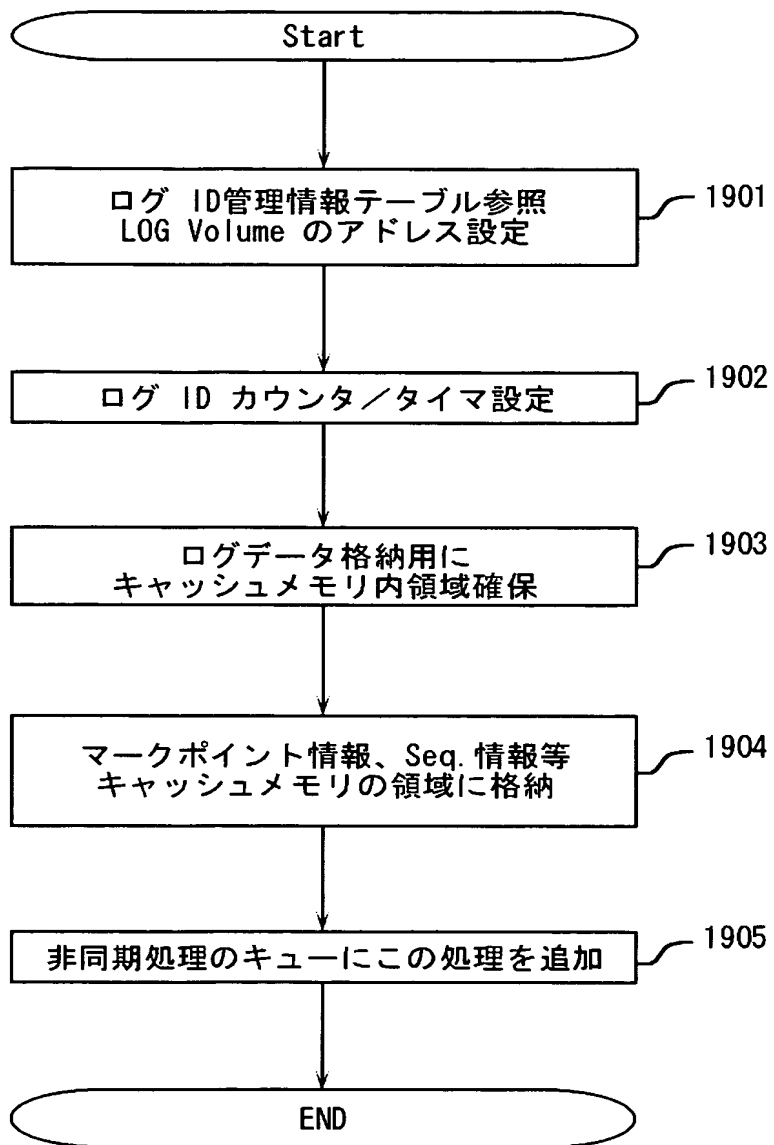
## コマンド処理（リモート側）



【図 20】

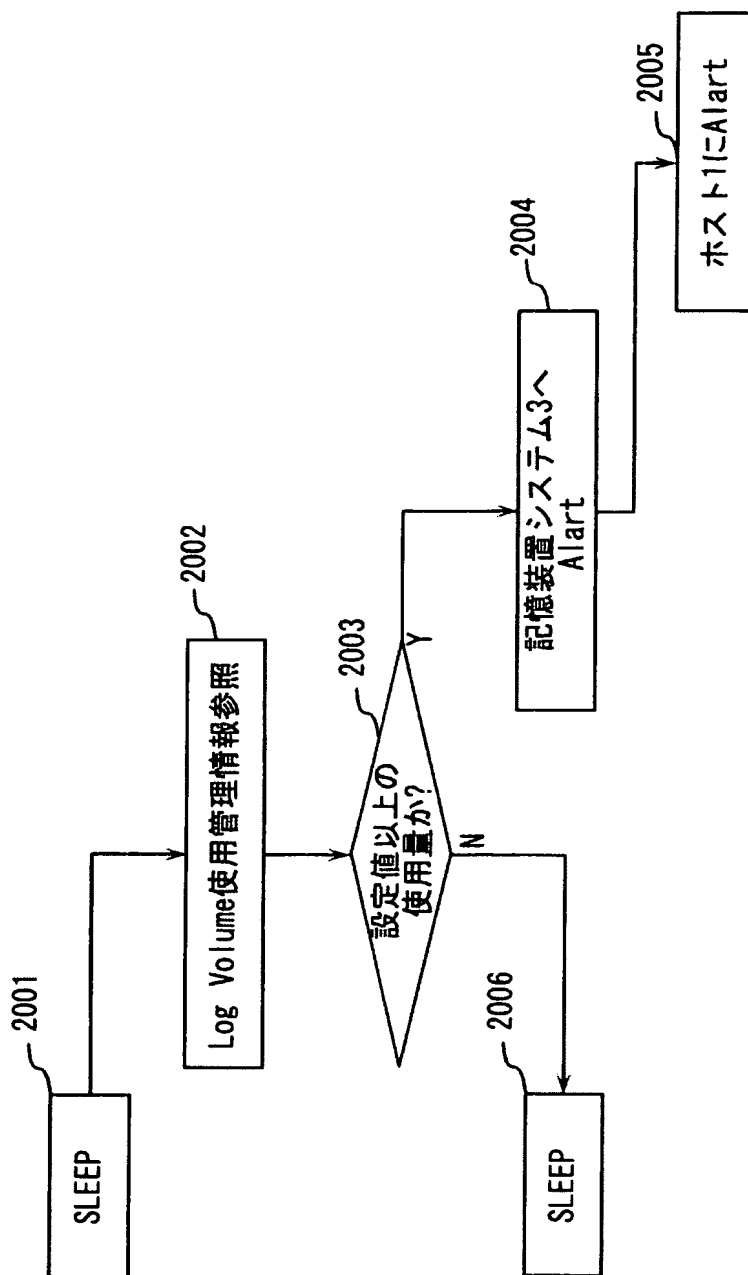
図 20

ログデータ作成処理（リモート側）



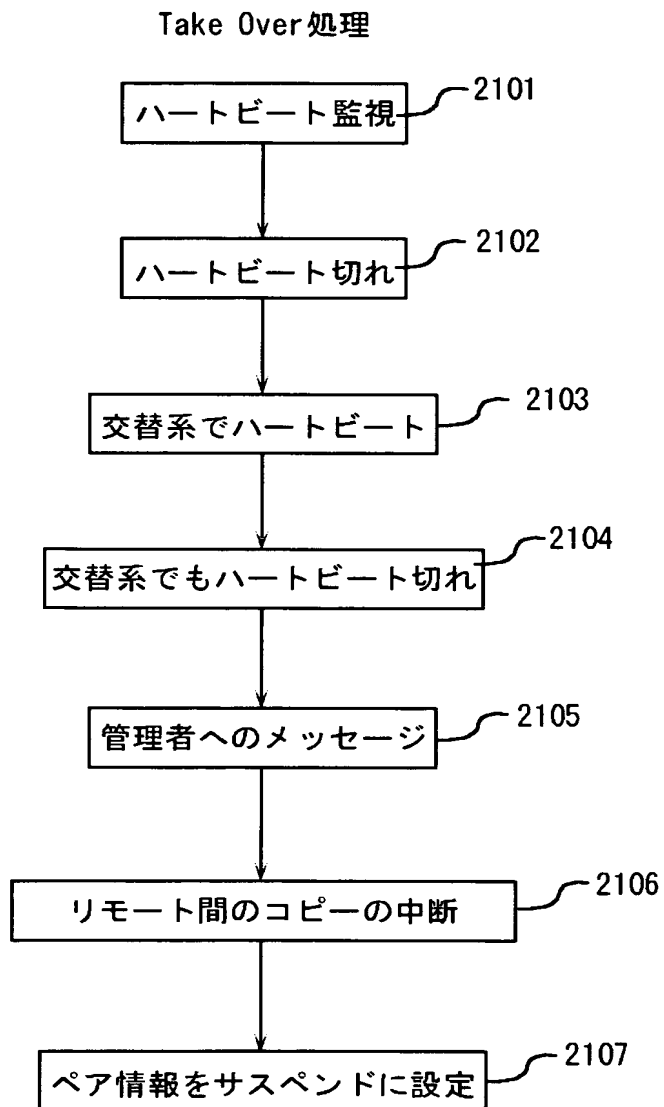
【図 21】

図 21  
Volume容量オーバー処理



【図 22】

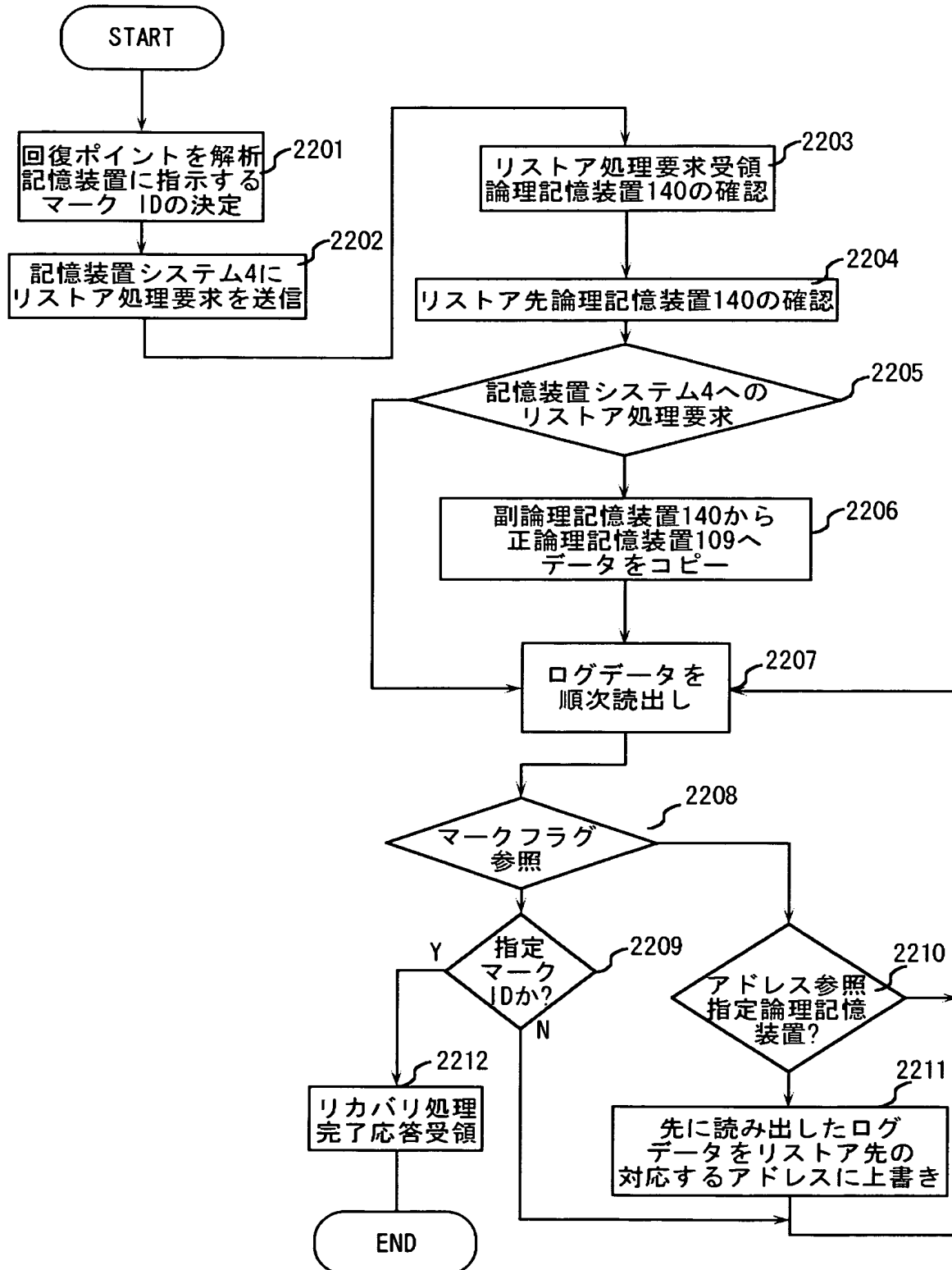
図 22



【図 23】

図 23

## Recovery処理



**【書類名】 要約書****【要約】****【課題】**

ホストに負担をかけず、障害発生時点までのデータをリモートサイトで高速にリストアする。

**【解決手段】**

マスタ側の第1の記憶装置システムは、ホストからの入出力要求を処理し、かつリモート側の第2の記憶装置システムに対して入出力処理の結果、更新されたデータを送信し、第2の記憶装置システムは、第1の記憶装置システムからの受信したデータを更新ログデータとして保持する。ホストはアプリケーションの状態確定するコマンドをデータとして第1の記憶装置システムに送信し、第1の記憶装置システムはこのデータを第2の記憶装置システムに送信する。かつホストと第2の記憶装置システムは、コマンドに対応した識別子を双方で保持し、識別子とログデータとを関連付けることにより、ホストが任意の時点で識別子を指示することによって第2の記憶装置システムで任意の時点のデータを復元する。

**【選択図】** 図2

特願 2 0 0 4 - 0 2 6 3 5 6

出 願 人 履 歴 情 報

識別番号 [ 0 0 0 0 0 5 1 0 8 ]

1. 変更年月日 1 9 9 0 年 8 月 3 1 日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台 4 丁目 6 番地

氏 名 株式会社日立製作所